



The Automatic Learning for the Rapid Classification of Events (ALeRCE) Alert Broker

F. Förster^{1,2,3}, G. Cabrera-Vives^{4,2}, E. Castillo-Navarrete^{1,2}, P. A. Estévez^{2,5}, P. Sánchez-Sáez^{2,6,7,8}, J. Arredondo²,
 F. E. Bauer^{6,9,2,10}, R. Carrasco-Davis^{2,5}, M. Catelan^{6,9,2}, F. Elorrieta^{2,11,20}, S. Eyheramendy^{7,2}, P. Huijse^{12,2},
 G. Pignata^{13,2}, E. Reyes^{2,5}, I. Reyes^{2,1,5}, D. Rodríguez-Mancini^{2,4}, D. Ruz-Mieres^{1,2,14}, C. Valenzuela^{21,7,2},
 I. Álvarez-Maldonado^{1,2}, N. Astorga^{2,5}, J. Borissova^{15,2}, A. Clocchiatti^{6,2}, D. De Cicco^{6,2}, C. Donoso-Oliva^{4,2},
 L. Hernández-García^{2,15}, M. J. Graham¹⁶, A. Jordán^{7,2}, R. Kurtev^{15,2}, A. Mahabal^{16,17}, J. C. Maureira¹,
 A. Muñoz-Arancibia^{2,1}, R. Molina-Ferreiro⁵, A. Moya^{2,1}, W. Palma², M. Pérez-Carrasco^{4,2}, P. Protopapas¹⁸, M. Romero⁵,
 L. Sabatini-Gacitua^{2,1}, A. Sánchez^{2,4}, J. San Martín¹, C. Sepúlveda-Cobo^{2,1}, E. Vera¹, and J. R. Vergara^{19,2}

¹ Center for Mathematical Modeling, University of Chile, AFB170001, Chile; francisco.forster@gmail.com

² Millennium Institute of Astrophysics, Nuncio Monseñor Sotero Sanz 100, Providencia, Santiago, Chile

³ Departamento de Astronomía, Universidad de Chile, Casilla 36D, Santiago, Chile

⁴ Department of Computer Science, University of Concepción, Chile

⁵ Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, Santiago 8320000, Chile

⁶ Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile

⁷ Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Av. Diagonal las Torres 2640, Peñalolén, Santiago, Chile

⁸ Inria Chile Research Center, Av. Apoquindo 2827, Las Condes, Chile

⁹ Centro de Astroingeniería, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile

¹⁰ Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, CO 80301, USA

¹¹ Department of Mathematics and Computer Science, Universidad de Santiago de Chile, Av. Libertador Bernardo O'Higgins 3663, Estación Central, Santiago, Chile

¹² Instituto de Informática, Universidad Austral de Chile, General Lagos 2086, Valdivia, Chile

¹³ Departamento de Ciencias Físicas, Universidad Andres Bello, Avda. Republica 252, Santiago, Chile

¹⁴ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

¹⁵ Instituto de Física y Astronomía, Universidad de Valparaíso, Av. Gran Bretaña 1111, Playa Ancha, Chile

¹⁶ Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA

¹⁷ Center for Data Driven Discovery, California Institute of Technology, Pasadena, CA 91125, USA

¹⁸ Institute for Applied Computational Science, Harvard University, Cambridge, MA, USA

¹⁹ Departamento de Informática y Computación, Universidad Tecnológica Metropolitana, Santiago, Chile

²⁰ Center for Interdisciplinary Research in Astrophysics and Space Exploration (CIRAS), Universidad de Santiago de Chile, Av. Libertador Bernardo O'Higgins 3363, Estación Central, Chile

²¹ Data Observatory Foundation, Santiago, Chile

Received 2020 June 29; revised 2021 February 22; accepted 2021 February 25; published 2021 April 27

Abstract

We introduce the Automatic Learning for the Rapid Classification of Events (ALeRCE) broker, an astronomical alert broker designed to provide a rapid and self-consistent classification of large etendue telescope alert streams, such as that provided by the Zwicky Transient Facility (ZTF) and, in the future, the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST). ALeRCE is a Chilean-led broker run by an interdisciplinary team of astronomers and engineers working to become intermediaries between survey and follow-up facilities. ALeRCE uses a pipeline that includes the real-time ingestion, aggregation, cross-matching, machine-learning (ML) classification, and visualization of the ZTF alert stream. We use two classifiers: a stamp-based classifier, designed for rapid classification, and a light curve-based classifier, which uses the multiband flux evolution to achieve a more refined classification. We describe in detail our pipeline, data products, tools, and services, which are made public for the community (see <https://alerce.science>). Since we began operating our real-time ML classification of the ZTF alert stream in early 2019, we have grown a large community of active users around the globe. We describe our results to date, including the real-time processing of 1.5×10^8 alerts, the stamp classification of 3.4×10^7 objects, the light-curve classification of 1.1×10^6 objects, the report of 6162 supernova candidates, and different experiments using LSST-like alert streams. Finally, we discuss the challenges ahead in going from a single stream of alerts such as ZTF to a multistream ecosystem dominated by LSST.

Unified Astronomy Thesaurus concepts: Supernovae (1668); Variable stars (1761); Active galactic nuclei (16); Astrominformatics (78); Surveys (1671); Classification (1907); Astrostatistics (1882); Convolutional neural networks (1938); Random Forests (1935); Cloud computing (1970); Distributed computing (1971); Small solar system bodies (1469)

1. Introduction

The exponential growth of the light-collecting area of telescopes and the number of pixels of digital detectors has resulted in a new generation of survey telescopes that are revolutionizing the way we study the time domain in astronomy (Tyson 2019). New surveys that systematically scan the optical/near-infrared sky with deep, wide, and fast-cadence observations (e.g., Catalina Real-Time Transient Survey, CRTS, Drake et al. 2009; Palomar Transient

Factory, PTF, Law et al. 2009; Optical Gravitational Lensing Experiment, OGLE, Udalski et al. 2015; Dark Energy Survey, the Dark Energy Survey Collaboration 2005; SkyMapper, Keller et al. 2007; Kepler, Koch et al. 2010; Vista Variables in the Via Lactea Survey, VVV, Minniti et al. 2010; Korea Microlensing Telescope Network, KMTNet, Kim et al. 2016; Hyper Suprime-Cam Subaru Strategic Program, HSC-SSP, Aihara et al. 2018; Asteroid Terrestrial-impact Last Alert System, ATLAS, Tonry et al. 2018;

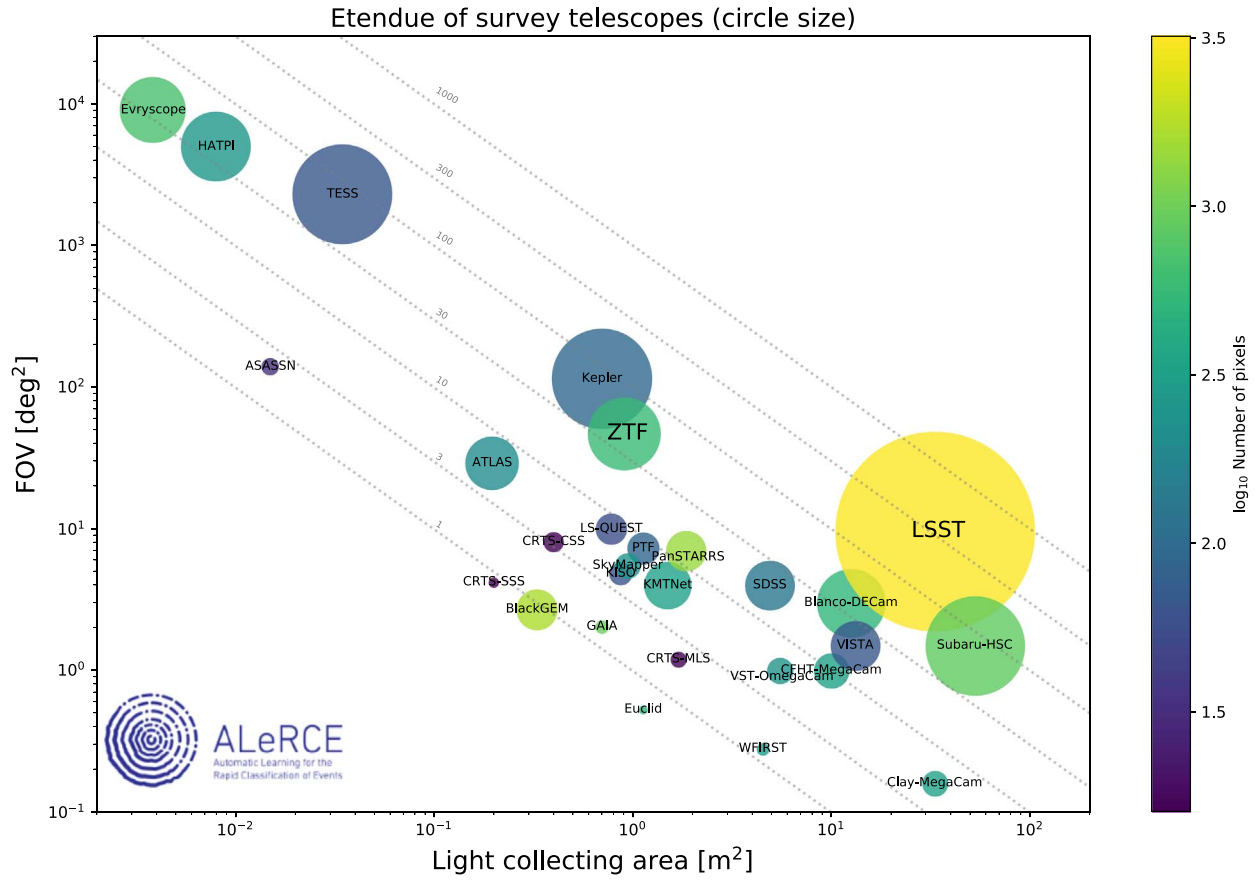


Figure 1. The FOV vs. light-collecting area for a selection of ground- and space-based survey telescopes currently operational or planned. The product of the two is called etendue and is indicated by the relative sizes. Note that if a survey contains several identical telescopes, we consider the sum of their etendues. The color of the circles indicates the number of pixels in the main camera of the instrument, following the color coding on the right. Constant etendue loci are shown as gray dashed lines, with the specific etendue value shown for each line. See Table 4 for telescope names and references.

Zwicky Transient Facility, ZTF, Bellm et al. 2019; Deeper, Wider, Faster, Andreoni et al. 2020) are uncovering large populations of time-varying astrophysical phenomena, including new populations of dim, rare, and/or short-lived events (e.g., Kasliwal et al. 2012; Drout et al. 2014).

As the construction of the Vera C. Rubin Observatory and its Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009) is advancing, a convergence with surveys in other regions of the electromagnetic spectrum (e.g., Square Kilometre Array, SKA, Dewdney et al. 2009; Wide-field Infrared Survey Explorer, WISE, Wright et al. 2010; eROSITA, Merloni et al. 2012; Fermi Gamma-ray Space Telescope, Atwood et al. 2009; Cerenkov Telescope Array, CTA, Actis et al. 2011), high-energy particles (e.g., CTA; IceCube Neutrino Observatory, Aartsen et al. 2017), and gravitational waves (Laser Interferometer Gravitational-Wave Observatory, Abramovici et al. 1992; Advanced Virgo, Acernese et al. 2015) is opening a new era of multimessenger astronomy (Abbott et al. 2017; IceCube Collaboration et al. 2018).

The fundamental quantity that defines a survey telescope is the product of mirror area and field of view (FOV), known as etendue, which is a simple proxy for the volume in space that can be monitored by different telescopes for the same exposure time and a given intrinsic luminosity object. We show the FOV, collecting area, and number of pixels of a selection of large etendue survey telescopes in Figure 1.

The detectors in these large etendue telescopes produce data at increasingly faster rates. Millions of events, i.e., objects that are seen to change their brightness or position in the sky, are being detected and reported in the form of continuous astronomical alert streams (Patterson et al. 2019). These streams create an opportunity for a new generation of follow-up telescopes to characterize large numbers of astronomical events in a coordinated fashion.

A new time-domain ecosystem is being built accordingly, where telescopes specialize as either survey or follow-up telescopes, but also new digital information components are developed to connect them seamlessly. The aggregation, annotation, and classification of alerts in a rapid and consistent fashion is done by astronomical alert brokers, such as the Automatic Learning for the Rapid Classification of Events (ALeRCE; this work); Alert Management, Photometry and Evaluation of Lightcurves (Nordin et al. 2019); Arizona-NOAO Temporal Analysis and Response to Events System (Narayan et al. 2018); Fink (Möller et al. 2021); LASAIR (Smith et al. 2019); and Make Alerts Really Simple.²² Different brokers typically specialize in different science cases. Their main role is to provide a fast and consistent classification of the alert stream using all of the available data but also enable filtering of the stream for different scientific communities. The fast classification of events is critical for the study of either short-lived phenomena or the early phases of evolution of

²² <https://Mars.lco.global/>

longer-lived processes, enabling follow-up observations to occur fast enough for some physical properties to be inferred (e.g., Gal-Yam et al. 2014). They will also contribute to the detection of new astrophysical phenomena in the form of outliers/anomalies (e.g., Nun et al. 2016) and will help reveal new subpopulations among known families of events (e.g., Baron & Poznanski 2017).

An interoperable and agile ecosystem is needed, with all of the relevant parts able to interact automatically to perform coordinated observations but also capable of adapting quickly to new science cases, instruments, or digital technologies. In this new scenario, follow-up telescopes will listen and react to Target and Observation Managers (TOMs; e.g., Street et al. 2018). TOMs will listen to alert broker classified streams, and brokers will listen to survey telescope alert streams. When follow-up observations are performed and their results become available, TOMs will be able to modify their follow-up strategy, brokers will be able to improve their classification, and survey telescopes will be able to change their surveying strategies, providing a feedback mechanism for the entire time-domain ecosystem to continuously improve.

1.1. Alert Broker Challenges

Astronomical alert brokers are a new kind of tool in the interface between astronomy and data science. They face new challenges, including infrastructure, machine-learning (ML), and community integration. This makes them important laboratories for testing new ideas in data science going even beyond astronomy.

In terms of infrastructure, the biggest challenge for astronomical brokers is to ingest, annotate, and classify, in a scalable fashion, the large astronomical alert streams coming from large etendue telescopes such as ZTF or LSST. For example, we have typically received between 10^5 and 10^6 alerts night⁻¹ from the public ZTF stream, associated with 5.1×10^7 objects as of 2021 February. For comparison, LSST is expected to produce about 10^7 alerts night⁻¹ and contain more than 10^9 different objects, which requires a distributed type of database and processing. Additionally, there will be a diversity of survey streaming alerts that must be cross-matched and classified in real time (e.g., ZTF, ATLAS, LSST). Thus, the challenge is to ingest data streams from a diversity of telescopes in a scalable fashion and classify them using their combined information to enable a rapid reaction by follow-up telescopes and a self-consistent analysis.

In terms of ML development, the challenges are diverse. What is an appropriate and relevant taxonomy for the astronomical community? How should we balance classification purity and efficiency? How can we develop ML classifiers and bring them into production in a reasonable timescale? How should we include cross-matched information in these classifiers? How can we train models using data that may be highly unbalanced and not fully representative of the unlabeled data? For example, training a classifier with spectroscopically labeled data will tend to be biased toward the bright end of the magnitude distribution. How can we train in a semisupervised fashion to take advantage of the unlabeled data? How can we train using data from a different telescope with a different set of filters/cadences (i.e., transfer learning and domain adaptation)? How can we train models using synthetic or augmented data? How can we detect outliers in a stream of data? All of these are technically challenging problems that need to be developed,

validated with the community, and then brought quickly into production.

Integration with the time-domain ecosystem and its community of users is another important challenge. First, brokers must be connected with other brokers, follow-up infrastructure, and data exploration tools. For this to happen, application programming interfaces (APIs) must be developed, simple interfaces that allow users to interact with different databases using virtual observatory or de facto standards. Second, in order to produce relevant data products and tools, frequent interaction with the community is needed to provide feedback and inject new ideas that can help improve the entire ecosystem. This includes interaction with small to large projects that interoperate with the community of survey telescopes, brokers, TOMs, and follow-up telescopes. A diversity of brokers must be encouraged, avoiding a winner-take-all solution and fostering an environment where new, creative solutions rise faster into production.

1.2. The ALeRCE Broker

The ALeRCE broker is a Chilean-led project that aims to become a community broker for LSST and other large etendue survey telescopes. The project is run by an interdisciplinary team composed of astronomers, computer scientists, and engineers, including faculty, postdoctoral fellows, and students. The broker's concept was first announced in 2017 as the natural continuation of the High cadence Transient Survey (HiTS), in which we used the Dark Energy Camera on the 4 m Blanco telescope to discover supernovae (SNe) in real time by combining tools from high-performance computing and ML (Förster et al. 2016). In 2018, a team of scientists was consolidated, the key requirements were defined, the first version of the front end was developed, a memorandum of understanding was signed with the ZTF project, and the initial funding was secured. In early 2019, a dedicated team of engineers was hired to start building the tools needed to ingest the public ZTF alert stream in preparation for LSST.

ALeRCE started to systematically classify the ZTF stream using ML with astrophysically motivated taxonomies based on their light curves (Sánchez-Sáez et al. 2021) in March 2019 and image stamps (Carrasco-Davis et al. 2020) in July 2019. These classifiers are designed to balance the need for a fast and simple classification with a subsequent but more complex classification. ALeRCE has reported 6162 SN candidates to the Transient Name Server²³ (TNS), of which 883 have been spectroscopically classified. It has classified 1.1×10^6 objects into a taxonomy that has expanded into 15 classes, including transient, periodic, and stochastic variable sources, and with continuously improving precision and purity. All of ALeRCE's data products can be accessed freely via several dashboards, APIs, or a direct database connection.

ALeRCE has adopted Agile work methodologies,²⁴ which have been adapted to academic environments by several groups.²⁵ Adopting this methodology has important implications for the broker, which becomes a continuously evolving product with regular data and code releases. All of the major components become dynamic: the classification taxonomy, as

²³ <https://wis-tns.weizmann.ac.il/>

²⁴ <https://agilemanifesto.org/>

²⁵ <https://www.agilealliance.org/resources/experience-reports/reinventing-research-agile-in-the-academic-laboratory/>

the available data sources grow and the product owners identify new scientific questions; the ML classification models, as new training sets and ideas are brought from development into production; and the tools and products, in order to adapt to the changing requirements of the community of users. This means that special attention needs to be given to version control of the broker pipeline, tools, and data products. This is done via the use of GitHub repositories to track code changes and the Semantic Versioning²⁶ naming convention for our pipeline and associated data releases.

The outline of this paper is as follows. In Section 2, we introduce the science goals of the ALerCE broker, including a discussion of the broker taxonomy. In Section 3, we describe the ML classifiers used by our broker. In Section 4, we present the pipeline structure and its associated infrastructure. In Section 5, we discuss our main data products, services, and tools. In Section 6, we present some of the main results. Finally, in Section 7, we draw some conclusions and discuss future directions.

2. Science Goals

Our primary science goals are the study of three broad categories of objects: transients, variable stars, and active galactic nuclei (AGNs). We also provide solar system object classifications as a secondary science goal.

2.1. Transients

Two important questions that can be answered via the study of transients are: (1) what is the nature of explosive phenomena, and (2) what can they teach us about the dynamics of the universe. Rapid classification is key to answering these questions, since it can facilitate dedicated follow-up observations, either rapid or slow, spectroscopic or photometric. Rapid follow-up is critical to understanding short-lived transients and the progenitors of stellar explosions in general, since it probes the outermost, unprocessed layers of exploding stars and the possible interaction with the circumstellar medium (e.g., Yaron et al. 2017; Förster et al. 2018). Early spectroscopy can be used to measure the composition and velocity structure of their ejecta. Late-time follow-up, either photometric or spectroscopic, probes the nature of the progenitor and explosion mechanism by constraining the composition and velocity structure of the innermost layers of the star (e.g., Fang et al. 2019). Having large samples of classified transient events cross-matched with multiband/messenger or contextual information will help characterize the parameter space and provide clues to new, unrecognized populations of events. Furthermore, the ability to cross-match different streams in real time, e.g., the LIGO and LSST streams, will offer possibilities that can lead to new, unexpected discoveries. These larger and better calibrated samples with well-understood systematics can be used for cosmological distance and/or event rate estimations. Finally, rapid follow-up of gravitational microlensing events can allow the detection of planets with masses and separations resembling those in our solar system (e.g., Bennett & Rhie 1996; Gould et al. 2010), while microlensing events with timescales of the order of years can provide clues about the nature of black holes (BHs) and dark matter (e.g., Green 2016). Moreover, microlensing may allow spectroscopic follow-up of sources that might otherwise have been too faint for spectroscopy (e.g., Bensby et al. 2020).

2.2. Variable Stars

Some of the important questions that can be answered via the study of variable stars are: (1) what is the nature of these systems and the physical mechanisms of variability, and (2) what can they teach us about the structure and formation of our own galaxy, its satellites, and other galaxies in the Local Group (e.g., Catelan & Smith 2015, and references therein). There are various reasons to obtain a uniform and rapid classification of variable stars. Rapid follow-up of stars entering/leaving the instability strip or changing their pulsation modes could provide new insights into the physics of stellar pulsation (e.g., Clement & Goranskij 1999; Buchler & Kolláth 2002; Soszyński et al. 2014). Detection and follow-up of eclipses in pulsating stars can help provide direct stellar mass measurements (e.g., Pietrzyński et al. 2010, 2012). The detection of eruptive events and the spectroscopic follow-up immediately after the beginning of the eruption can provide new insights into the physics of young stellar objects (Contreras Peña et al. 2017; Connelley & Reipurth 2018). Finally, larger and more distant samples of consistently classified variable stars (e.g., Gaia Collaboration et al. 2019a) will be key to understanding the tridimensional structure and formation history of our galaxy, along with that of its neighbors, ranging from the ultrafaint dwarfs to the Magellanic Clouds (e.g., Dékány et al. 2019; Jacyszyn-Dobrzeńska et al. 2020a, 2020b; Vivas et al. 2020).

2.3. Active Galactic Nuclei

Some of the most exciting questions that can be answered from the study of AGN are: (1) what drives the growth of BHs (Alexander & Hickox 2012), (2) what are the physical mechanisms behind AGN variability (Ross et al. 2018; Sánchez-Sáez et al. 2018), (3) are there intermediate-mass BHs (IMBHs; Mezcua 2017; Greene et al. 2020) with masses between stellar and supermassive BHs (SMBHs), (4) what is the structure and size of AGNs (Lawrence 2016), and (5) what can tidal disruption events (TDEs; Arcavi et al. 2014) teach us about BH properties. Rapid classification could help identify and follow up on optical changing-look AGNs, a population that may unlock numerous clues to BH accretion physics (LaMassa et al. 2015; Graham et al. 2020). Selecting large samples of targets based on their multiband variability for reverberation mapping studies can enable better physical constraints on the BH surrounding medium and distance (Peterson et al. 2004). Fast-cadence data can help assemble large samples of IMBH candidates (Martínez-Palomera et al. 2020), which are known to vary on shorter timescales. The early detection of TDEs can provide independent constraints on the BH properties that drive these phenomena (Komossa 2015). All of the above can be done while simultaneously cross-matching the LSST stream with future surveys that will provide critical additional information, such as eROSITA (Merloni et al. 2012), SKA precursors, IceCube (Abbasi et al. 2009), etc. Finally, exploring larger samples of AGNs that are dimmer and redder can lead to the discovery of new populations of events and a better understanding of the AGN phenomena.

3. ML Classification

3.1. Classification Taxonomy

An important component of an automatic classifier is the taxonomy used for classification, which defines the classes into

²⁶ <https://semver.org/>

Table 1
Light Curve–based ML Classifiers that Include Both Transient and Persistent Variable Objects

Reference	Data Source	Data Type	No. of Classes	Classes
Sánchez-Sáez et al. (2021) (See Section 3.3)	ZTF	Observed	15	SN Ia, SN Ibc, SN II, SLSN, AGN, QSO, blazar, CV/nova, YSO, DSCT, RRL, Ceph, LPV, E, periodic–other
Boone (2019)	PLAsTiCC	Simulated	14	AGN, RRL, E, Mira, M dwarf, ML, TDE, kN, SN Ia, SN Ia-91bg, SN Iax, SN Ibc, SN II, SLSN-I
Martínez-Palomera et al. (2018)	HiTS	Observed	8	NV, QSO, CV, SN, DSCT, E, ROT, RRL
Narayan et al. (2018)	OGLE, OSC	Observed	7	SN, BPer, RRL, LPV, Ceph, DSCT, DPV
D’Isanto et al. (2016)	CRTS	Observed	6	CV, SN, blazar, AGN, M dwarf, RRL

Note. Note that Sánchez-Sáez et al. (2021) is an accompanying publication where we describe the ALeRCE light-curve classifier in more detail.

which the alert stream will be classified. Choosing a good taxonomy is about achieving a balance between a reasonably accurate classifier, which depends on finding good training sets and the intrinsic separability of the classes, and meeting the demands of different communities of users. More complex taxonomies can be useful for a larger set of communities, but the addition of subclasses can lead to potentially less accurate classification models. The best compromise between the accuracy of the classifier and the complexity of the taxonomy is difficult to define; therefore, in order to guide our choice of taxonomy, we performed a survey of the taxonomies used in other studies that carried out ML classification of variable astronomical objects.

3.1.1. Light-curve Classifier Taxonomy

First, we consider those works that use only light curves in their analysis. We divide them into those that include both persistent variable and transient sources (Table 1), those that include only persistent variable objects (Tables 5 and 6), and those that include only transient objects (Table 7). We examined four publications that include both transient and persistent variable objects in their taxonomy, 22 publications that include only persistent variable objects, and eight publications that include only transient objects. There were 19 different sources of observational data, mostly for persistent variable sources (Table 8), and five sources of synthetic data (Table 9).

A large diversity of taxonomies was found, with fewer classes in general being used in the last 5 yr with respect to older works. This may be due to the appearance of more exploratory efforts in recent years, which look for variations from more traditional classification methods while using fewer classes for simplicity. We found more classes of persistent variable objects of stellar origin, probably because of the relative abundance of curated light-curve training sets for these classes. The synthetic data sources were applied mostly for transient data, probably because of the relative difficulty in finding large numbers of observed transients. A brief description of the classes is included in the Appendix. The pulsating star variable classes included in the previous publications are shown in Tables 10 and 11, other stellar variable sources in Table 12, SMBH-related sources in Table 13, and transients in Table 14.

In general, there are certain families of objects that seem to be included consistently among most classifiers but whose decomposition into subclasses varies greatly. Taking this into account, we have decided to develop a hierarchical classifier

that groups families of classes and will gradually be refined as the amount and quality of the data grow (Sánchez-Sáez et al. 2021). The first level of the classifier considers transient, periodic, and stochastic variable phenomena. In the second level, the transient branch divides into (class name abbreviations in parentheses) the Type Ia SN (SN Ia), Type Ib and Ic SN (SN Ibc), Type II and IIn SN (SN II), and superluminous SN (SLSN) classes. The periodic branch divides into the eclipsing binary (E), δ Scuti (DSCT), RR Lyrae (RRL), Cepheid (Ceph), long-period variables (LPVs, including Miras and semiregular and irregular variables), and other (periodic–other) classes. The periodic–other class corresponds to periodic objects that are not members of the E, DSCT, RRL, Ceph, or LPV classes. The stochastic branch divides into host-dominated AGNs, core-dominated AGNs, quasi-stellar objects (QSOs), blazars, cataclysmic variables and novae (CV/nova), and young stellar objects (YSOs).

ALeRCE’s current classification taxonomy is shown in Figure 2. This figure draws inspiration from the variability diagram of Eyer & Mowlavi (2008), most recently updated in Gaia Collaboration et al. (2019b) but significantly simplified and with a more observationally based hierarchy, more resolution in the transient classes, and less resolution in the stellar variability classes. The reason for having more resolution in the transient classes is that in many cases, the reaction time for the photometric or spectroscopic follow-up of these classes needs to be fast, e.g., to get spectroscopic confirmation or characterize a short-lived phase of evolution, while for the persistent variability classes, it is not as common to require fast follow-up. Thus, our main goal is to provide a first filter for the expert communities to explore further and classify into more complex taxonomies in more branches of the classification tree.

3.1.2. Stamp Classifier Taxonomy

In addition to the classifiers that work solely on light curves, there are classifiers that use the pixel information contained on the variable object detection images. Alerts are generated from a difference image that results from aligning, scaling, convolving, and subtracting the reference image from the science image. We have listed the ML classification studies that use the object “image stamps” in Table 2 for the classification of images into either real or bogus (e.g., Bloom et al. 2012) but also as members of more astrophysically motivated classes. The latter efforts are relevant for the taxonomy of our stamp-based classifier, a classification model that uses as input the

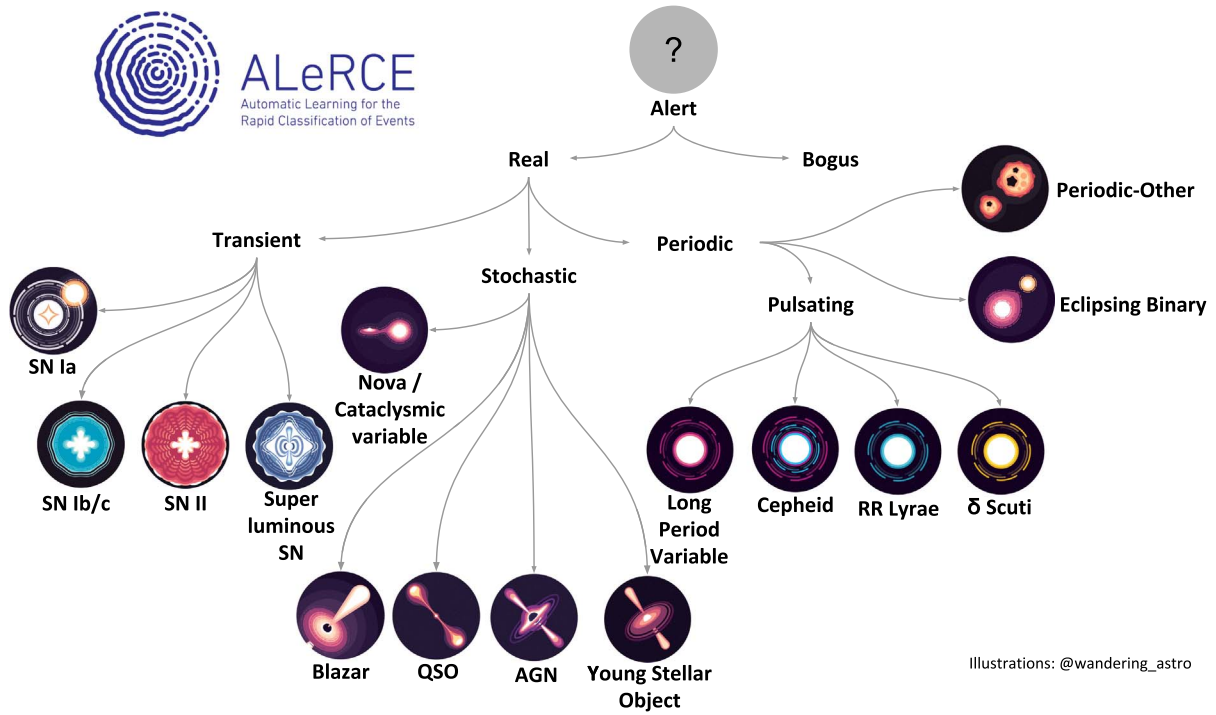


Figure 2. Hierarchical taxonomy used by the ALerCE broker for classifying light curves (v1.0.0). This classifier uses four models: one that separates transient, stochastic, and periodic objects; another that separates transients into SNe Ia, SNe Ib/c, SNe II, and SLSNe; another that separates stochastic objects into blazars, QSOs, AGNs, and YSOs; and another that classifies periodic stars into LPVs, Ceph, RRL, DSCT, E, or periodic-other.

Table 2
Single Image Stamp ML Classifiers

Reference	Data Source	No. of Classes	Classes
Carrasco-Davis et al. (2020) (Section 3.4)	ZTF	5	SN, AGN, VS, asteroid, bogus
Duev et al. (2019)	ZTF	2	Real, bogus
Wright et al. (2017)	PanSTARRS1	3	Real, asteroid, bogus
Cabrera-Vives et al. (2017)	HiTS	2	Real, bogus
Kimura et al. (2017)	HSC-SSP	2	SN Ia, other
du Buisson et al. (2015)	SDSS	2	Real, bogus
Carrasco et al. (2015)	RCS-2	2	Stars, QSOs
Bloom et al. (2012)	PTF	5	Bogus, suspect, unclear, maybe, realish
Bailey et al. (2007)	PTF	2	Real, bogus

Note. Empirical data are used in all cases. Note that Carrasco-Davis et al. (2020) is an accompanying work where we describe the ALerCE stamp classifier in more detail.

first set of science, template, and difference images associated with a new object in the alert stream²⁷ and is used as the first classification step in ALerCE. Although the complexity of the taxonomy associated with this classifier is less refined, this early classification is critical to enable the triggering of fast photometric and spectroscopic follow-up and characterization of extragalactic transient sources. In the case of our stamp-based classifier (Carrasco-Davis et al. 2020), we have used the classes SN, AGN, variable star (VS), asteroid, and bogus, trying to mimic how astronomers have historically looked for transients and variables. The SNe tend to be near extended sources; AGNs are either relatively isolated pointlike sources or at the center of extended sources, depending on luminosity; variable stars are pointlike sources that are frequently near other pointlike sources and present in both the science and

reference images; asteroids are present only in the science image, not in the reference image; and bogus sources are not shaped like the point-spread function (PSF) of the image.

Finally, we found one publication that uses time series of image stamps (Carrasco-Davis et al. 2019) following an approach that combines time series and image stamps using a convolutional recurrent neural network classifier. They use seven classes: nonvariable, galaxy, asteroid, SN, RRL, Ceph, and E. This type of work could become more important in the future because it combines spatial and temporal information, as well as simulated and real data.

3.2. Training Sets

In order to compile training sets, we use only sources observed by ZTF whose labels have been cross-matched from different catalogs available in the literature or compiled by our collaboration. For each catalog, we define a function that maps the

²⁷ Note that the same object can have many associated alerts.

catalog’s taxonomy into our own taxonomy, allowing us to aggregate labels from different catalogs into a unified taxonomy. Then, we assign a priority order that defines which labels to use in case of disagreement between catalogs. These priorities are based on discussions with community experts, a critical analysis of the methods that were used to classify objects (e.g., manual versus automatic), and an analysis of which catalogs tend to disagree more with other catalogs from a visual exploration of catalog label matrices (similar to confusion matrices but with rows and columns as the classes in each catalog, potentially with different taxonomies).

The catalogs we use to extract labels are, in order of priority,

1. the cataclysmic variables catalog, compiled by Abril et al. (2020), including Ritter & Kolb (2003);
2. ROMABZCAT, the multifrequency catalog of blazars from Massaro et al. (2015);
3. the catalog of type I AGNs from Oh et al. (2015);
4. the Million Quasars Catalogue from Flesch (2019);
5. the spectroscopically classified SNe in the TNS;²⁸
6. the objects classified as YSOs in Simbad (Wenger et al. 2000);
7. the CRTS catalog of northern periodic sources (Drake et al. 2014);
8. the CRTS catalog of southern periodic sources (Drake et al. 2017);
9. the LINEAR catalog of periodic variables (Palaversa et al. 2013);
10. the Gaia Data Release 2 (DR2) catalog of variable stars (Mowlavi et al. 2018); and
11. the All-Sky Automated Survey for Supernova (ASAS-SN) catalog of variable stars (Jayasinghe et al. 2019).

3.3. The Light-curve Classifier

This classifier computes classification probabilities for objects with ≥ 6 detections in g or r . We represent individual light curves as a vector of features compiled from the literature and new features developed by the ALERCE collaboration as described in Sánchez-Sáez et al. (2021). One of the most relevant new features comes from an irregularly sampled autoregressive model (IAR) introduced in Eyheramendy et al. (2018), which is able to estimate autocorrelation in irregularly sampled time series in a statistically robust way. The classification is done in a hierarchical fashion using a balanced random forest classifier,²⁹ which, in our tests, achieved better accuracies than recurrent neural networks. As described before, a given object will be first classified as either periodic, stochastic, or transient and subsequently refined into 15 different classes as described in Section 3.1. The confusion matrix associated with this classifier can be seen in Figure 3, described in Sánchez-Sáez et al. (2021).

3.4. The Stamp Classifier

Inspection of ZTF image stamps suggests that it should be possible to classify alerts based on the first detection set of stamps (see Section 3.1.2). Therefore, we designed and trained a stamp classifier based on a convolutional neural network with the main motivation of finding SN candidates, using as input

the information contained in the first alert, including the science, reference, and difference stamp set, as well as other metadata, such as spatial location and data quality metrics.

The stamp classifier (Carrasco-Davis et al. 2020) is able to discriminate among five classes, SNe, AGNs, variable stars, asteroids, and bogus alerts, achieving 90% accuracy on a balanced test set and a recall of 81% among spectroscopically confirmed SNe from TNS. To improve the model interpretability, we added a regularization term that maximizes the entropy of the predicted probability for each class, enhancing the different certainties for each prediction. This model is currently running on ZTF alerts, and its results are publicly available in the ALERCE SN Hunter at <https://snhunter.alerce.online> (see Section 5.2.1). The confusion matrix associated with this classifier can be seen in Figure 4, reproduced from Carrasco-Davis et al. (2020).

3.5. Metrics and Selection of Classification Model

In order to evaluate the classifiers that will go from initial model training into production, we use a combination of metrics and tests that take into account the labeled and unlabeled data. We have found this to be relevant when using a labeled training set known to be nonrepresentative of the unlabeled data. First, we compute the test set classification balanced (averaged per class) accuracy (ratio between correct and total labels) and F1 score (the harmonic mean between precision and recall) to take into account the accuracy, precision, and recall of the classifier while considering the class imbalance, which is very important when using observational data as training sets. Second, we look at the confusion matrix to search for signs of overrepresentation of certain classes that may not be evident in the balanced accuracy. Third, for the light-curve classifier, we look for classification biases with certain relevant variables, e.g., looking for a relatively constant recall versus apparent magnitude relation for individual classes when no significant bias exists. Fourth, we compare the expected and inferred spatial and class distributions of the unlabeled data to discard models using astrophysical knowledge. For example, if the classification model were correct, one would expect the spatial distribution of the different classes to follow known patterns, such as that most Galactic classes should be concentrated around the Galactic plane, extragalactic classes should be homogeneously distributed outside the Galactic plane due to extinction and source confusion, and asteroids should be distributed around the ecliptic. Additionally, we would expect the distribution of class labels in the unlabeled set to follow known population ratios; for example, we expect SNe Ia to be more abundant than SNe Ibc. Therefore, the final choice of a classification model is made considering all of these metrics and tests before the model is brought into production, i.e., applying the model using the available infrastructure with our latest pipeline for nightly operations.

3.6. Stamp and Light-curve Classifier Comparison

As a consistency check between the two aforementioned classifiers, we compare the distribution of classes of the stamp classifier among those objects classified by the light-curve classifier. In Figure 5, we show a matrix of stamp and light-curve classifier classes, normalized along the light-curve classifier classes. We can see that there is overall agreement between the two

²⁸ <https://wis-tns.weizmann.ac.il/>

²⁹ Using the `imblearn` library.

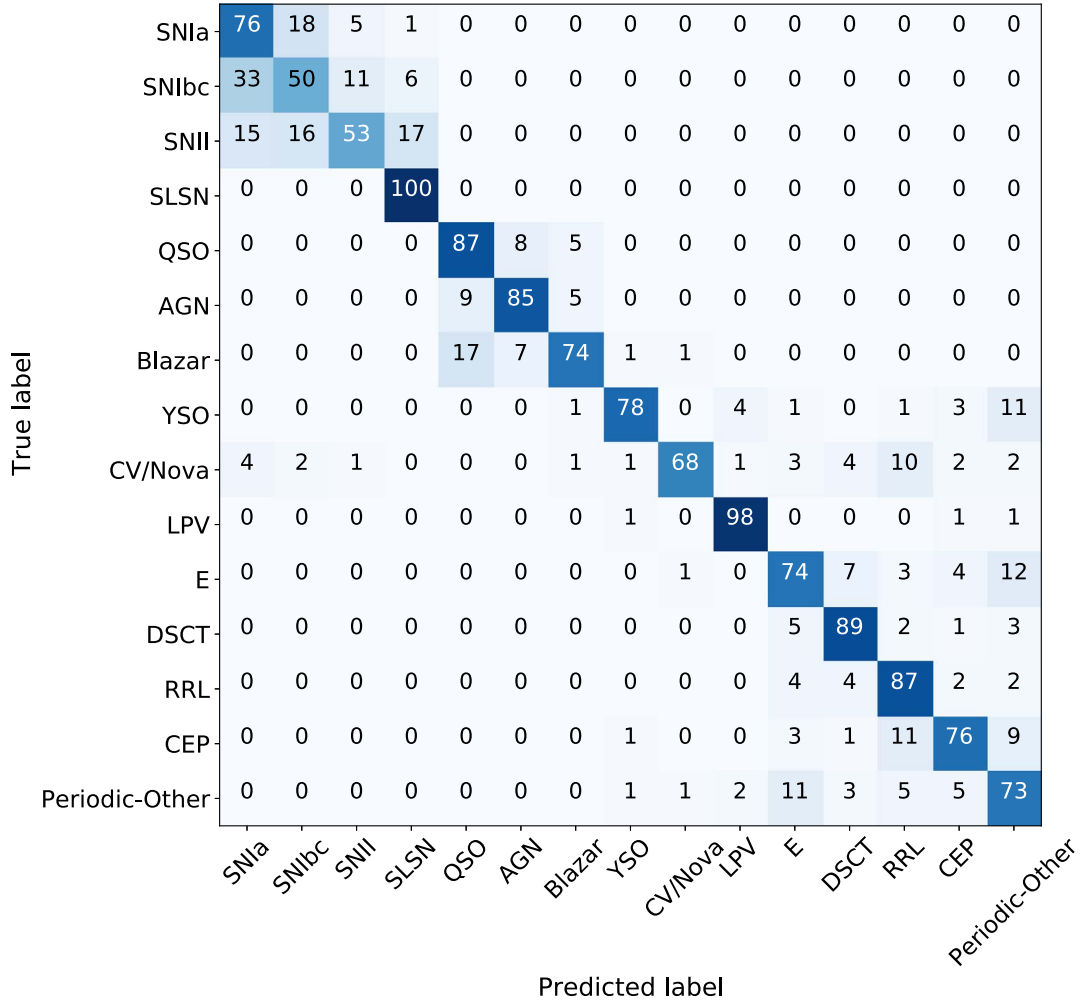


Figure 3. Confusion matrix obtained with the balanced hierarchical random forest light-curve classifier model in Sánchez-Sáez et al. (2021).

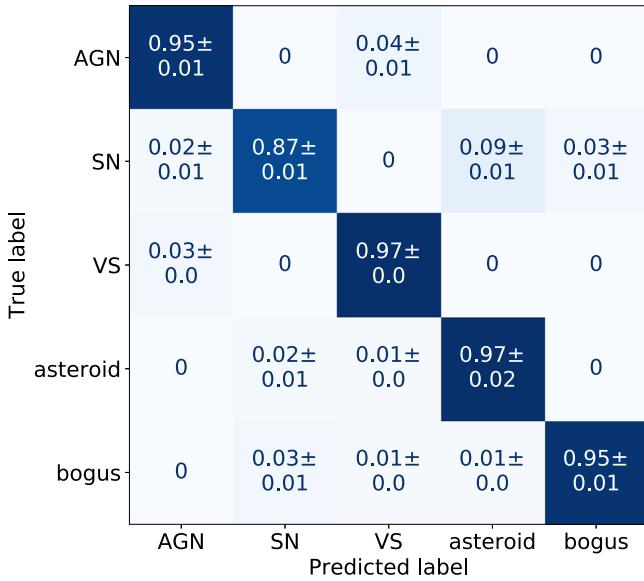


Figure 4. Confusion matrix obtained with the stamp classifier model in Carrasco-Davis et al. (2020).

classifiers, which highlights the complementarity between our two classifiers and emphasizes the value of using the image stamps for early classifications, as shown in Carrasco-Davis et al. (2019).

3.7. Outlier/Novelty Detection

Outlier/novelty detection refers to the automatic identification of abnormal or unexpected phenomena embedded in data (Faria et al. 2016). We are developing outlier detection methods experimentally to focus on two problems: the discrimination of outlier clusters of time series or image stamps, i.e., cohesive and representative sets of examples associated with interesting phenomena that are not characterized in the current training database, and the detection of unexpected events occurring within a particular time series. To solve the first problem, we are developing online one-class/semisupervised outlier detection methods (Schölkopf et al. 2001; Chapelle et al. 2009; Reyes & Estévez 2020) to find similarities between objects and automatically detect outlier phenomena. We are addressing this problem from three different perspectives: using autoencoders, generative adversarial networks, and one-class neural networks. To find unexpected events within time series, we are using robust online nonlinear filters (Liu et al. 2011; Huentelemu et al. 2016). Traditional methods, such as Kalman and kernel filters, are being extended to incorporate measurement uncertainties, the heteroscedasticity of the noise, and the use of state space formulations where states are unevenly separated in time.

For both problems, active-learning techniques (Zhu et al. 2003) are being explored to select sets of the most uncertain objects and/

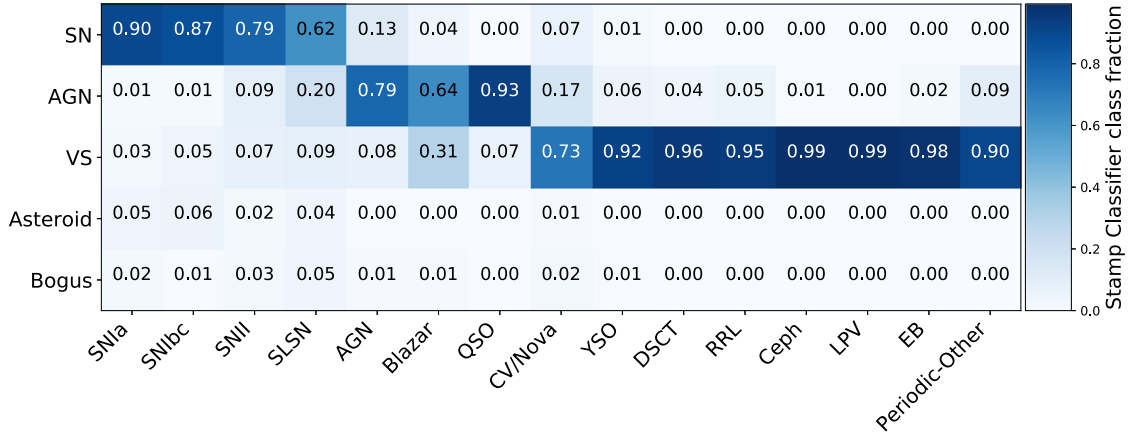


Figure 5. Fraction of objects predicted to belong to a given stamp classifier class (rows), normalized among the objects predicted to belong to a given light-curve classifier class (columns). We considered a sample of 186,794 unlabeled objects that were classified with the stamp (Carrasco-Davis et al. 2020) and light-curve (Sánchez-Sáez et al. 2021) classifiers.

or events to be shown to human experts. We are aiming to use information theoretical feature selection (Estévez et al. 2009) and feature extraction methods to reduce dimensionality and generate visualizations that can be presented to the experts.

4. ALerCE Pipeline and Infrastructure

ALerCE is currently processing the alert stream provided by the ZTF survey, but we expect to ingest other alert streams in the future, such as those provided by ATLAS, HATPi,³⁰ and LSST (see Figure 1). The ZTF pipeline and alert distribution system are described in Masci et al. (2019) and Patterson et al. (2019). Alert packets contain image difference stamps and other metadata, whose detailed description can be found at <https://zwickytransientfacility.github.io/ztf-avro-alert/schema.html>. The ALerCE system ingests these alerts and processes them through a pipeline that is divided into a combination of sequential and parallel steps, shown schematically in Figure 6 and described below.

4.1. Ingestion and Kafka Topics

The ZTF alerts are sent as Avro packets,³¹ a data serialization format that contains associated image stamps, metadata, and information related to previous detections as described in <https://zwickytransientfacility.github.io/ztf-avro-alert/schema.html>. We use Apache Kafka,³² a framework for working with streaming data, to receive the ZTF alert stream and communicate information between the different steps of our pipeline as independent Kafka topics. We use an Apache Zookeeper cluster with a replication factor of 3, following recommended practices, and three independent machines of Kafka consumers that are responsible for reading data from the alert queue. We have set up a Kafka cluster in Amazon Web Services (AWS) to manage different topics associated with different steps in the pipeline. Assigning different topics for each step in the pipeline has the advantage of allowing for alerts to be grouped in different batch sizes optimized for performance. For example, querying the database for several objects simultaneously can be faster than doing it sequentially for a list of objects depending on the type of query, or, in the case of cross-matching, it may be more efficient to group alerts

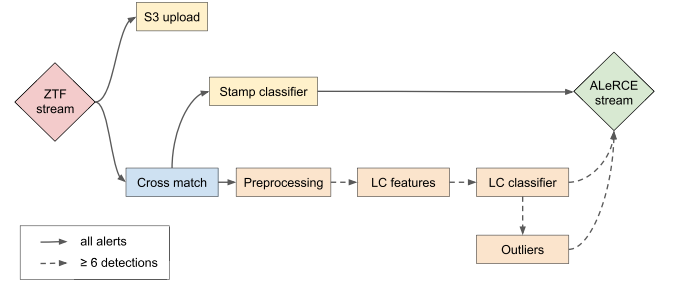


Figure 6. ALerCE pipeline structure from ZTF alert ingestion to the ALerCE streaming of the processed alert. Alerts ingested from the public ZTF stream are first sent to four parallel Kafka topics: an Avro backup service in AWS S3, the stamp classifier for early SN detections, a cross-match step to gather information from public catalogs, and a light curve (LC) correction step. The LC correction step is followed by an LC features computation step and LC classifier and outlier detection steps, which are only applied to objects with six or more detections. Note that the ML classification steps can also be fed with information from the cross-match step. The tables of our database are modified inside the pipeline steps for subsequent access via APIs.

by their spatial location if the external catalog is stored hierarchically, e.g., a tessellation of the sky. Another advantage is that we can configure each topic independently for performance, e.g., using different numbers of Kafka partitions per topic.

We have tested different configurations of Kafka producers to mimic an LSST-like stream of data, and we have found that a cluster of three Kafka consumers with 12 partitions each is capable of ingesting all of the different topics at a rate of 119.7 MB s^{-1} , that is, about three times faster than the average alert production rate expected for LSST.

4.2. Database and Avro Repository

As alerts arrive, we store the original Avro files in AWS Simple Storage Service (S3) buckets for future analysis and extract a selection (in order to limit the size of the database) of the fields contained in these packets to be added directly to a database using a PostgreSQL database engine. As the data are processed and object alerts aggregated, we add different statistics to different tables. The main tables in our database are as follows.

³⁰ <https://hatpi.org/science/>

³¹ <https://avro.apache.org>

³² <https://kafka.apache.org>

1. The `object` table, which contains basic filter and time-aggregated statistics, such as location, number of observations, and times of first and last detection.
2. The `magstats` table, which contains time-aggregated statistics separated by filter, such as the average magnitude or initial magnitude change rate.
3. The `detection` table, which contains the object light curves, including their difference and corrected magnitudes and associated errors separated by filter (see Section 4.4).
4. The `non_detection` table, which contains the limiting magnitudes of previous nondetections separated by filter.
5. The `feature` table, which contains the object light-curve statistics and other features used for ML classification that are stored as json files in our database.
6. The `xmatch` table, which contains the object cross-matches and associated cross-match catalogs.
7. The `probability` table, which contains the object classification probabilities, including those from the stamp and light-curve classifiers and different versions of these classifiers.
8. The `taxonomy` table, which contains details about the different taxonomies used in our stamp and light-curve classifiers, which can evolve with time.

A webpage containing an updated description of the different tables can be found at <https://alerce.science>. As the volume of alerts grows for different projects, we expect to migrate some of the previous tables to NoSQL database engines such as Cassandra or MongoDB. After ingestion, the alerts undergo the processing steps described next.

4.3. Stamp Classification

When an alert from a previously unreported object arrives, its first available image stamps are used to classify it as either SN, AGN, variable star, asteroid, or bogus, as explained in Section 3.4. Note that if the first detection from an object did not pass the ZTF real/bogus test but a subsequent detection did, the first available image stamp will not be from the former. This stamp classification is done within 1 s of the alert being received and is automatically available in our database and the SN Hunter tool (see Section 5.2.1), if the candidate is consistent with being an SN. The details of the stamp classifier are described in a parallel publication (Carrasco-Davis et al. 2020).

4.4. Light-curve Correction

As explained before, ZTF alerts are produced when a science image contains a significant change with respect to a reference image after aligning, scaling, convolving, and subtracting the reference image from the science image. Flux differences with respect to the reference image are reported as difference magnitudes, and an associated flag (`isdiffpos`) is included to indicate whether the difference is positive or negative. In the case of ZTF, a reference image is defined by a unique reference field identifier (`rfid`). If the source was present in the reference image, it is possible to recover its actual apparent magnitude from the difference and reference magnitudes. We do this correction when the nearest cataloged object is closer than $1''.4$ (`distnr < 1.4`), providing a flag to indicate whether we think the object is extended based on PanSTARRS and ZTF shape parameters. The actual apparent magnitude, m_{corr} , and associated

errors, δm_{corr} , in the case of a pointlike source that was present in the reference are the following:

$$m_{\text{corr}} = -2.5 \log_{10}(10^{-0.4 m_{\text{ref}}} + \text{sgn } 10^{-0.4 m_{\text{diff}}}), \quad (1)$$

$$\delta m_{\text{corr}} = \frac{(10^{-0.8 m_{\text{diff}}} \delta m_{\text{diff}}^2 [-10^{-0.8 m_{\text{ref}}} \delta m_{\text{ref}}^2])^{0.5}}{10^{-0.4 m_{\text{ref}}} + \text{sgn } 10^{-0.4 m_{\text{diff}}}}, \quad (2)$$

where m_{ref} is the magnitude of the object in the reference image, m_{diff} is the magnitude associated with the absolute flux difference between the science and reference images, `sgn` is the sign of the difference (`isdiffpos`), δm_{ref} is the error associated with the reference magnitude, and δm_{diff} is the error associated with the difference magnitude. Note that we provide both the original and corrected photometry. For the corrected photometry, we include errors values with and without the term inside square brackets in Equation (2) that originates from the correlation between the reference and difference fluxes (see derivation in Appendix).

It is important to note that if the difference flux is equal to the reference flux and the sign of the difference is negative, both the corrected magnitude and associated errors will diverge, which is a limitation of using a logarithmic scale for difference fluxes. This should normally not occur, since an alert is triggered only when there is a significant difference with respect to the reference. However, if the reference image contains a transient source, the difference flux can eventually become exactly minus the reference flux, and the corrected flux is zero, which will lead to divergences depending on the noise. We treat these cases by assigning values of 100 to the corrected magnitudes and their associated errors.

We discuss in detail the derivation of these formulae, how to include the effect of a change in the reference image, and how we treat extended sources in the reference image in Appendix.

4.5. Cross-match

A cross-match step runs with the stamp classifier and light-curve correction, querying external catalogs in order to extract additional information about the objects of interest. The ZTF alert packets already contain the nearest solar system, PanSTARRS, and Gaia cataloged sources. In addition to this information, we query WISE and SDSS in order to obtain infrared and spectroscopic information, if available, which can be critical to better constraining some of the classes included in our taxonomy. Additional catalogs will be included as they prove relevant. These queries are done using the CDS cross-match API,³³ which can handle sufficiently large streams if alerts are grouped in batches around the same region of the sky before querying.

4.6. Feature Computation

With the corrected light curves, we can compute light-curve characteristics or features based on both the detections and nondetections of a given object, as well as available cross-matches. Advanced light-curve features are only triggered for objects with ≥ 6 detections in *g* or *r*. The features computed are a significantly extended version of the FATS library (Nun et al. 2017), called Turbo FATS, that is optimized for computation

³³ <http://cdsxmatch.u-strasbg.fr/xmatch/doc/>

Table 3
Summary of ALerCE Data Products and Services as of 2021 February

Type	Name	Address
Database	ALerCE DB PostgreSQL repository	db.alerce.online
GitHub repositories	ALerCE open-source repositories	http://github.com/alercebroker
Jupyter notebooks	Science use-case notebooks	http://github.com/alercebroker/usecases
Jupyter notebooks	TNS upload notebooks	http://github.com/alercebroker/TNS_upload
Output stream	ALerCE output Kafka stream	Please contact us.
Website	ALerCE main website	http://alerce.science/
Website	ALerCE workshops website	http://workshops.alerce.online/
Dashboard	ALerCE Grafana pipeline dashboard ^a	http://grafana.alerce.online/
Documentation	ALerCE API documentation	http://alerceapi.readthedocs.io/en/latest/
Documentation	ALerCE client documentation	http://alerce.readthedocs.io/en/latest/
Documentation	ALerCE tutorial videos	https://bit.ly/2NHDagc
Web interface	ALerCE explorer	http://alerce.online
Web interface	SN Hunter	http://snhunter.alerce.online
Web interface	Cross-match interface	http://xmatch.alerce.online
Web interface	ALerCE reporter	http://reporter.alerce.online/
Web interface	TOM Toolkit plug-in	http://tom.alerce.online/
API	ZTF DB access	http://ztf.alerce.online
API	Avro/stamp service	http://avro.alerce.online
API	ZTF cross-match service	http://xmatch-api.alerce.online
API	catsHTM cross-match service	http://catshrm.alerce.online
API	TNS cross-match service	http://tns.alerce.online
API	Finding chart generator	http://findingchart.alerce.online

Note.

^a Request access.

speed and adds several new features. A description of these features, which are contained in the `feature` table of our database, can be found in Sánchez-Sáez et al. (2021).

4.7. Light-curve Classification

Objects having computed features are then processed by the light-curve classifier described in Section 3.3. The results of this classifier are obtained within a few tens of seconds from ingestion for 95% of the objects. For a larger stream, this could be maintained by scaling the infrastructure given the embarrassingly parallel nature (i.e., no need of communication between parallel tasks) of the light-curve correction, feature computation, and light-curve classification tasks between different alerts. The current model used for the light-curve classifier is a hierarchical balanced random forest, as described in Sánchez-Sáez et al. (2021).

After the light-curve classification step, we perform an outlier detection step, which, as of 2021 February, is being actively developed experimentally (see Section 3.7).

4.8. Database Integrity Tests

After the nightly ingestion and processing of the alerts, we perform a series of database integrity tests during the day. This consists of reanalyzing the Kafka topic associated with the last night of observations to check that no alerts were lost during the processing due to unexpected errors. If any alerts were missed during the night, we add them to a specially created

Kafka topic that is then processed by our pipeline until no missing alerts exist.

5. Data Products and Services

The ALerCE broker provides several data products and services that are constantly growing as we identify new requirements from our community of users. New requirements are defined by user stories, informal descriptions of desired features from the perspective of an end user, that are translated into different data products and services by astronomers on our team following an Agile methodology. In this section, we list the most important data products and services provided by ALerCE as of 2021 February, which are summarized in Table 3.

5.1. Data Products

The ALerCE data products can be divided into several categories: the tables of a database, a repository of Avro files, a repository of Jupyter notebooks, an output stream of annotated and classified alerts, a GitHub repository with our open-source code, a Grafana dashboard to monitor the status of the pipeline, our main website, documentation websites, and tutorial videos for new users. We provide a brief description of each of them in what follows.

5.1.1. Database

The tables in our database integrate the information about individual objects. A description of the database can be found

in Section 4.2. The tables from our database are open for direct exploration in read-only mode, as shown in some of our use-case Jupyter notebooks (<https://github.com/alercbroker/usecases>), although we recommend accessing them using our different APIs for simple queries (see Section 5.2.2). A detailed description of the tables and schema used in our database can be found in <https://bit.ly/3oxhpbz>.

5.1.2. Avro Repository

Apart from the previous tables, a copy of the original Avro files contained in the ZTF stream are stored in AWS S3. These Avro files can be accessed using our Avro/stamp API.

5.1.3. GitHub Repositories

All of our open-source code can be found in the GitHub repository, <https://github.com/alercbroker>. In the course of developing this project and as of 2021 February, we have created 151 repositories, 42 of which have been made public for our community of users. These repositories can be forked or modified for external use. The pipeline steps are contained in these repositories, and new version numbers are defined when dockerized versions of the steps are created.

5.1.4. Use-case Jupyter Notebooks

We have compiled a list of example Jupyter notebooks that show how to use our API or directly access our database focused around different science cases, such as SN, variable star, AGN, and asteroid studies. They can be found at <https://github.com/alercbroker/usecases>.

Apart from these notebooks, we have created a special notebook and associated GitHub repository for the inspection and submission of SN candidates to TNS (https://github.com/alercbroker/TNS_upload). In this notebook, users can interact with Hierarchical Progressive Surveys (HiPS; Fernique et al. 2015) PanSTARRS images to easily select the candidate host galaxies using *ipyaladin*, *NED*, *Simbad*, and *SDSS DR15*. This repository includes a tutorial explaining all of the steps required to upload candidates to TNS, including tutorial videos to guide users in the process.

5.1.5. Output Stream

A real-time output stream is provided to report database changes as new alerts arrive and are processed by our pipeline, including an update on the classification probabilities and basic statistics. Users can connect to this stream using Apache Kafka upon request.

5.1.6. Grafana Dashboard

A Grafana dashboard is available to monitor the ALerCE pipeline and associated database and infrastructure (<http://grafana.alerce.online>). This dashboard shows the status of the Apache Kafka servers and relevant metrics about the number of alerts being processed, the PostgreSQL database and associated servers, and the front-end servers. Access to this dashboard can be given upon request.

5.1.7. Main Website, Documentation, and Tutorial Videos

ALerCE's main website, which summarizes all of our data products and services, can be accessed at <http://alerce.science>.

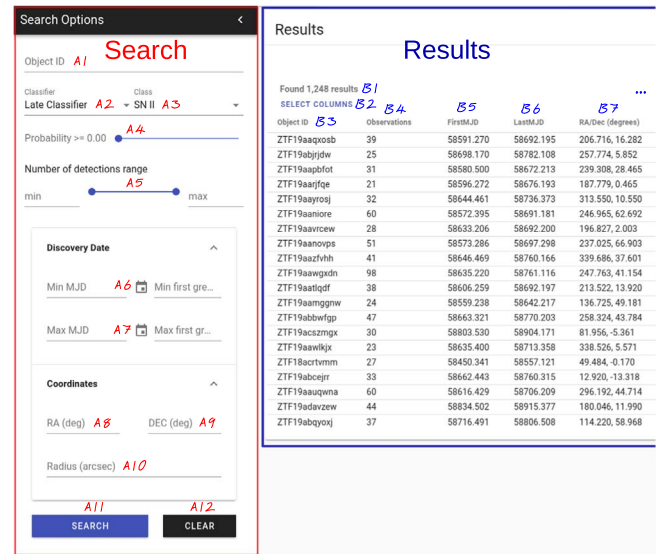


Figure 7. ALerCE explorer web interface (<http://alerce.online>) initial Search and Results view. The Search panel allows users to directly filter by object identifier (A1); inferred type using either the stamp or light-curve classification models (A2), a given class (A3), and a minimum classification probability (A4); minimum and maximum number of detections (A5); and minimum (A6) and/or maximum (A7) discovery date in Modified Julian Dates or calendar dates or by location in the sky using a cone search defined by an R.A. (A8), decl. (A9), and search radius (A10). The Search button (A11) submits queries, and the Clear button (A12) clears the search options. The Results panel shows the results of the previous query. First, it shows the total number of results (B1), which are displayed in a paginated format. Users can select which columns to display (B2). The columns shown in this figure are the object identifier (B3), number of detections (B4), time of first (B5) and last (B6) detection, and coordinates (B7). Other columns displayed by default (not shown in this image) are whether the object has cross-matches and the stamp and light-curve classifier classes and probabilities. Clicking on an object links to the Object view (Figure 8).

Documentation for our API services and client (see Section 5.2.1) and a series of detailed tutorial videos for our community of users can be found at <https://bit.ly/2YoEKbU>. A special website for workshops organized or coorganized by ALerCE is also available, with links to presentations and Jupyter notebooks that can be run using [Google Colab](https://colab.research.google.com/) for easier adoption.

5.2. Services

Apart from the previous data products, several services are provided to facilitate the exploration of the ZTF stream and associated objects. They are divided into web interfaces, which are websites that allow the simple exploration of the alert stream, and APIs, which power the previous web interfaces and allow for the flexible integration of ALerCE into the time-domain ecosystem.

5.2.1. Web Interfaces

ALerCE Explorer (<http://alerce.online>). The ALerCE explorer is the main tool to explore the astronomical objects recovered from the ZTF alert stream. Its landing page consists of two main sections: the Search and Results sections (see Figure 7). The Search section is where users can filter objects by selecting their unique identifier or different combinations of classifier, class, class probability, number of detections, and sky coordinates. The Results section is where the results of the filtered objects are shown, sorted by classification probability

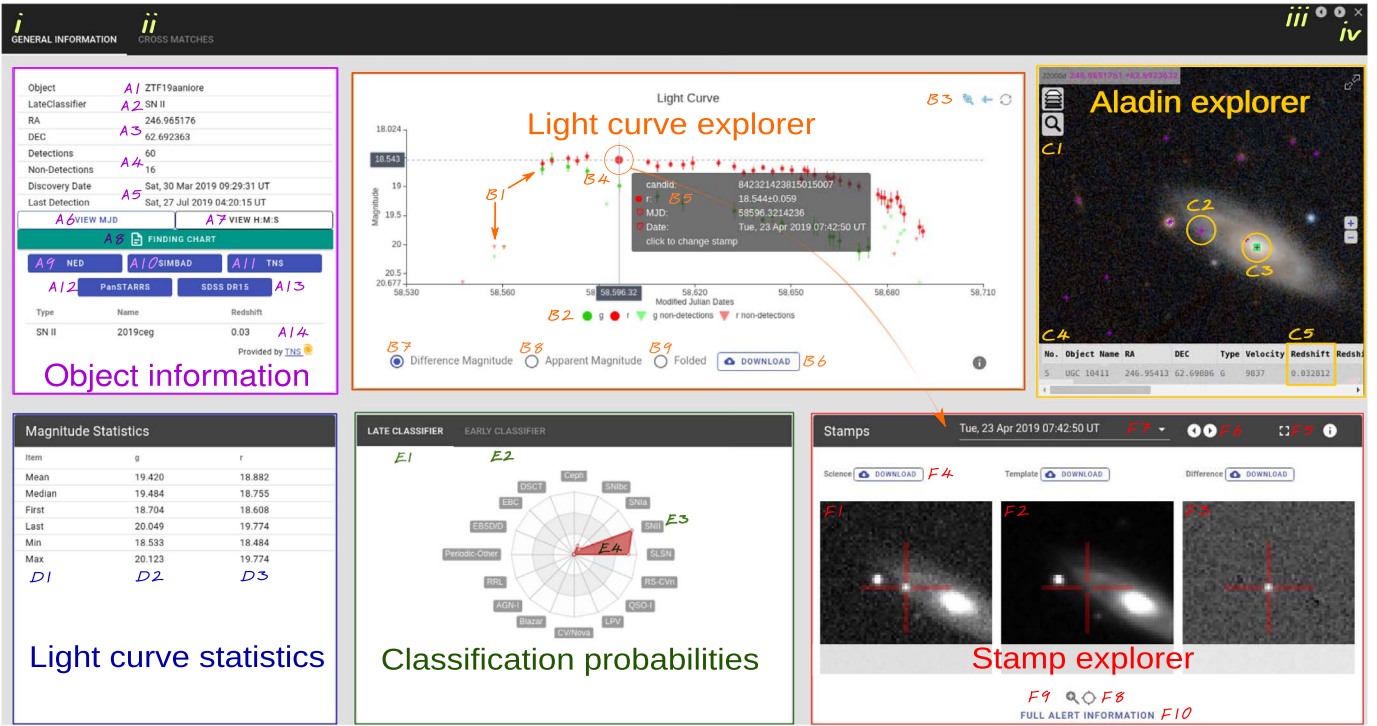


Figure 8. ALerCE explorer web interface (<http://alerce.online>) object view as of 2021 February. At the top left, users can switch between the General Information (i; this figure) or Cross Matches (ii) object views, and at the top right, between different objects (iii or the arrow keys) if directed from the results table of a previous query or to go back to the Search and Results view (iv or the escape key). The General Information view contains six different panels that we demarcate with colored text. At the top left, the object information panel shows the object’s unique identifier (A1), most likely class (A2), coordinates (A3) in different formats (A7), number of detections and nondetections (A4), and first and last detection times (A5) in calendar or Modified Julian Dates (A6). It also contains links to the finding chart generator tool (A8), NED (A9), Simbad (A10), TNS (A11), PanSTARRS (A12), and the SDSS DR15 navigation tool (A13). The latest type, name, and redshift associated with the object in the TNS are also shown (A14). At the top middle, the light-curve explorer panel displays the latest light curve of the object, including both detections and nondetection upper limits in both bands (B1), which can be turned on/off individually (B2). The light curve can be zoomed in and out (B3), and users can hover over individual points (B4) to see the exact date, magnitude, and alert identifier (B5) or click to display its associated stamps and full alert information in the stamp explorer panel. The light curve and associated data can be downloaded (B6), and users can select whether to show the difference magnitude (B7); the apparent magnitude (B8), i.e., corrected by the flux in the reference image; or the period-folded apparent magnitude (B9), assuming that the light curve is periodic and using a periodogram to compute the most likely period. At the top right, the Aladin explorer panel shows an interactive Aladin window (C1) with a PanSTARRS image at the location of the candidate (C2), in this case, a confirmed SN near its likely host galaxy (C3). An overlaid catalog of objects can be clicked on to view more information (C4), such as the host redshift (C5). At the bottom left, the light-curve statistics panel shows different statistics (D1) computed over the *g* (D2) and *r* (D3) bands of the apparent magnitude light curve. In the bottom middle, the classification probabilities panel shows the classification probabilities according to our light-curve (E1) or stamp (E2) classifiers, when available. A radar plot of the class probabilities for the taxonomy used in the classification model (E3) is shown. Hovering over the radar plot displays the numerical values of the probabilities (E4). At the bottom right, the stamp explorer panel shows the science (F1), reference (F2), and difference (F3) image stamps associated with any point in the light curve, which can be downloaded for further analysis (F4) or displayed in full-screen mode (F5). Users can switch between the previous or next stamps in time (F6) or select any particular date (F7) of the light curve that is contained in the public ZTF stream. Users can select between displaying crosshairs (F8) or simultaneously hovering and zooming (F9) over the stamps. They can also see the full alert information in the associated alert packet (F10). Note that those points in the light curve that do not pass the ZTF’s real/bogus test will not have stamps available for display, since they do not trigger an alert in the public stream.

or other variables. Clicking on an individual object will take the user to the object view page (see Figure 8).

The object view page is divided into two tabs, the General Information and Cross Matches tabs, with different panels each (see Figure 8). In the General Information tab, users can see some basic statistics about the object, generate a finding chart, query different catalogs at the position of the object (the NASA Extragalactic Database (NED), Simbad, TNS, PanSTARRS, or SDSS), or quickly see basic TNS information about the object. The user can see the object’s light curve, including detections and nondetections, with the capability of plotting the raw difference light curve, a corrected apparent magnitude (which includes the contribution of the reference image), or a folded version of the corrected apparent magnitude using the best-fitting period. The light-curve information can be downloaded as comma-separated values (CSVs), and every point in the light curve can be hovered over to see more information or clicked on to show its associated image stamp. HIPS images and

catalogs around the position of the object are shown using Aladin (Bonnarel et al. 2000), with superimposed NED and Simbad clickable objects. The science, reference, and difference image stamps associated with any point in the light curve can be shown in the stamps section, where the stamps can be explored by selecting different dates or hovering over them, seen in full screen, or downloaded as fits files. The full Avro packet information can also be explored. The classification probabilities are shown in the stamp and light-curve classifier tabs, where a radar plot is used to show the class probabilities assigned by the light curve– or stamp-based classifiers, if available. Finally, in the Cross Matches tab, users can see all of the cross-matches contained in the catsHTML set of catalogs for a given separation, which can be selected manually with a sliding bar (see Figure 9).

The ALerCE explorer is where most of our web development has been focused, including new tools, as requested by our community of users, but also new sources of data that in the

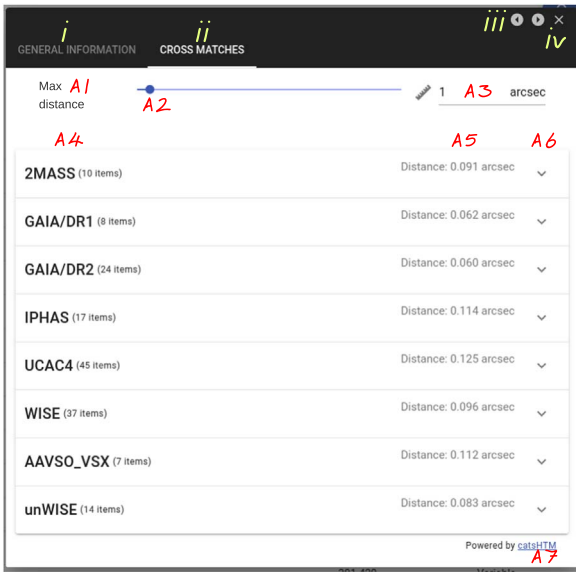


Figure 9. Object cross-matches view of the ALeRCE explorer as of 2021 February. Labels *i*, *ii*, *iii*, and *iv* are the same as in Figure 8. This panel allows users to find the closest cross-matching sources in the catsHTM data set, given a maximum cross-matching distance (A1) defined via a sliding bar (A2) or directly via its numeric value (A3). The closest cross-matches among different catalogs (A4) are shown with their associated distances (A5), allowing for an expanded view of the columns available in each catalog (A6). For more information, see the catsHTM (A7) reference (Soumagnac & Ofek 2018).

future will allow for the multistream exploration of astrophysical objects. We are developing a modular data exploration library that will be gradually expanded to include new sources of streaming data.³⁴ This library is used for a new version of the ALeRCE explorer that was being tested as of 2021 February, which is connected to a new database and can be previewed in <http://stage.alerce.online>.

SN Hunter (<https://snhunter.alerce.online>). The SN Hunter platform allows users to visualize and explore the best and most recent SN candidates (see Figure 10). These candidates are obtained using the convolutional neural network that powers the ALeRCE stamp classifier and can be seen in the SN Hunter just seconds after being received from ZTF. Users can see the spatial distribution of the candidates in celestial coordinates and in comparison to the Milky Way plane or the ecliptic, as well as a table that shows them sorted by classification probability, discovery date, or number of observations. Selecting a candidate displays an Aladin HiPS image at the location of the object, as well as the science, reference, and difference images contained in the Avro file. The candidates’s unique identifier, coordinates, and first observation properties and the properties of the closest PanSTARRS object are also shown, as well as links to the ALeRCE explorer for the same object or for NED, TNS, and Simbad sources around the position of the object. Users can also see the full alert information contained in the original Avro file of the alert by clicking the Full Alert Information button.

A key feature of the SN Hunter is the ability to receive feedback from users who have logged in. If a candidate appears to be bogus, users can label the candidate as such to further enhance the training set. Moreover, if the candidate appears to be an SN or extragalactic transient, the user can label it as a

possible SN to be sent to the ALeRCE reporter tool (see below). The list of possible SNe can then be explored by the team with our reporter tool, which can then be used to submit targets to the TOMs for follow-up. We regularly compute user labeling metrics in order to provide feedback or identify potentially malicious labeling, and we can use more than one label per object in order to prevent accidental labeling.

Reporter (<https://reporter.alerce.online>). The ALeRCE reporter tool is a platform that serves to manage user feedback in general (see Figure 11). As of 2021 February, it served three purposes: to manage the feedback provided by the SN Hunter interface, to connect with the TOM Toolkit interface, and to manage internal data classification challenges. The user feedback provided via the SN Hunter consists of bogus alert labels for alerts that appear to be bogus and possible SN alert labels for alerts or groups of alerts that appear to be originated by extragalactic transients. The connection of SN candidates with the TOM Toolkit interface is also done from the reporter tool, sending users to the TOM Toolkit interface after clicking on a reported candidate. Finally, the reporter tool can be used to create data challenges, manage associated user entries, produce metrics and confusion matrices, and show leader boards as in Kaggle. The data challenges are key for the collaboration’s periodic hackathons, where we set different classification challenges that motivate the ML team to develop new ideas and tools.

TOM Toolkit Plug-in (<https://tom.alerce.online>). This platform is used to manage and submit candidates to the TOM Toolkit (<https://lco.global/tomtoolkit/>). Users that have access rights to the ALeRCE reporter can connect with the TOM Toolkit via this interface, allowing them to submit observational requests with detailed instrumental specifications to the queue of different observatories.

Cross-match Service (<http://xmatch.alerce.online>). ALeRCE provides a cross-match service that allows users to submit an arbitrary CSV file with objects and coordinates of their favorite targets (see Figure 12). After a file is uploaded, the user is asked to select the names of the identifier, R.A., and decl. columns. After this is done, the closest objects in ZTF are returned, adding several columns from the ALeRCE object table to the submitted objects. A paginated table is shown for exploration, and the output can be downloaded as a CSV file.

5.2.2. APIs

All of the interactions between the web interfaces and the database or the Avro/stamp repository are done via APIs. These APIs serve most of ALeRCE’s data exploration tools following the principle of maximizing the modularization of our different services. They are also the key elements that will allow ALeRCE to integrate seamlessly with the astronomical time-domain ecosystem. These APIs are documented on the ALeRCE API Documentation website: <https://alerceapi.readthedocs.io/en/latest/>. Note that as of 2021 February, a new API was being tested that is better documented, easier to use, and connected to an entirely redesigned database (preview at <http://dev.api.alerce.online/>). Here we describe the services available as of 2021 February.

ZTF Database Access Service (<http://ztf.alerce.online>). This service allows users to query the ALeRCE database tables without needing any authentication. This API includes services to query objects filtered by unique object identifier, number of detections, class, class probabilities, coordinates, or detection

³⁴ <https://vue-components.alerce.online/>

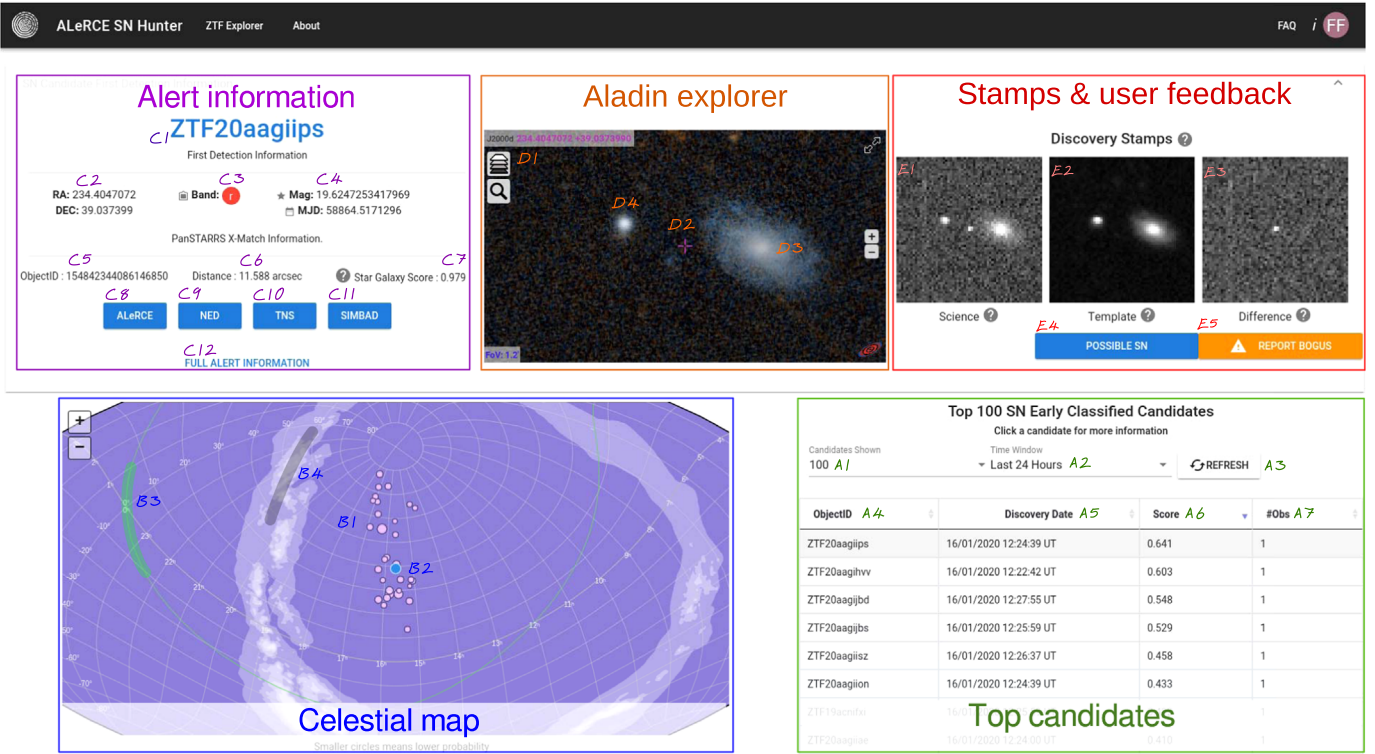


Figure 10. The SN Hunter web interface (<http://snhunter.alerce.online>) as of 2021 February, which allows users to find the highest stamp classification probability and most recent SN candidates in the ZTF alert stream in real time. This tool is divided into five panels and used by our collaboration to select candidates for submission to TNS. Starting at the bottom right, the top candidates panel shows a list of the top 10–1000 (default 100; A1) SN candidates in terms of their stamp classifier SN probabilities within the last 1–7 days (default 24 hr; A2). This list can be refreshed at any moment (A3). The results are shown in a paginated table sorted by either object identifier (A4), discovery date (A5), score or stamp classifier SN probability (A6), or number of detections (A7). Each candidate can be clicked on for exploration, opening up the top panels. At the bottom left, the celestial map panel shows the spatial distribution of all of the candidates in the top candidates panel, with a circle size proportional to their score (B1) and centered around the currently selected candidate (B2). Also shown are the position of the ecliptic (B3) and Milky Way plane, where the white contour levels crudely denote the density distribution of Galactic stars (B4). At the top left, the alert information panel shows the information about the currently selected candidate, including its object identifier (C1); coordinates (C2); band (C3); magnitude and time (C4) at first detection; and information about the closest PanSTARRS source, including its identifier (C5), distance (C6), and star galaxy score (C7, varying between zero and 1 between galaxies and stars). Links to the ALerCE explorer object view (C8), NED (C9), TNS (C10), and Simbad (C11) are provided. All additional information contained in the alert is also available for exploration (C12). At the middle top, the Aladin explorer panel provides an interactive Aladin window (D1) centered around the selected candidate (D2), where a host galaxy may be seen in PanSTARRS DR1 *gri* color images (D3). Note that although there is a clear host galaxy associated with this candidate, its closest source is a star (D4), which explains the star galaxy score displayed in C7. Finally, at the top right, the stamps and user feedback panel is where the science (E1), reference (E2), and difference (E3) ZTF image stamps are displayed for the currently selected candidate. If users are logged in using a Google account (i), they can label candidates as possible SNe (E4) or report them as bogus (E5) in order to improve the stamp classifier training set.

times. Users can also get the associated SQL command for a given query, all detections for a given object, all nondetections for a given object, the classification probabilities for a given object, or the features used as input for the ML classifiers for a given object. The documentation can be found at https://alerceapi.readthedocs.io/en/latest/ztf_db.html. This service is used in the ALerCE explorer and the SN Hunter (see Section 5.2.1).

Avro/Stamps Service (<http://avro.alerce.online>). This service allows users to access the alert Avro files and their associated stamps. The input is the unique object identifier and the unique stamp identifier. Users can get the Avro file, a specific field from an Avro file, or the science, reference, and difference image stamps contained in an Avro file. The documentation can be found at <https://alerceapi.readthedocs.io/en/latest/avro.html>. This service is used in the ALerCE explorer and the SN Hunter (see Section 5.2.1).

ZTF Cross-match Service (<http://xmatch-api.alerce.online>). This service allows users to submit an arbitrary catalog and get the nearest ZTF sources and their separation and properties. It is used in the cross-match interface (see Section 5.2.1).

catsHTM Cross-match Service (<http://catshmt.alerce.online>).

This service allows users to do cone searches to a given location using the catsHTM catalogs (Soumagnac & Ofek 2018). This includes cone searches returning all objects closer than a given distance from all or a specific catalog or only the closest object from all or a given catalog. This service is used in the ALerCE explorer Cross Matches view (see Section 5.2.1). The documentation, also indicating a list of all available catalogs, can be found at <https://alerceapi.readthedocs.io/en/latest/catshmt.html>.

TNS Cross-match Service (<http://tms.alerce.online>). This service allows users to query TNS information about an object centered around a given position in the sky. It queries the TNS API and returns the TNS name, type, and redshift, and it is used by the ALerCE explorer General Information tab (see Section 5.2.1).

Finding Chart Service (<http://findingchart.alerce.online>). This service provides a finding chart associated with a given object's unique identifier. It returns a PDF file with a PanSTARRS reference image indicating the location of the candidate, as well as the science, reference, and difference image stamps. An example finding chart can be seen in Figure 13. This service is used in the ALerCE explorer (see Section 5.2.1).

Candidate	Reports	Source(s)	First date reported	Last date reported	Action
ZTF20aasobhn	3	SN Hunter	Thu, 05 Mar 2020 14:58:06 GMT	Wed, 11 Mar 2020 15:21:28 GMT	
ZTF20aaszhtk	3	SN Hunter	Fri, 06 Mar 2020 14:05:21 GMT	Wed, 11 Mar 2020 15:21:22 GMT	
ZTF20aasxrko	3	SN Hunter	Sat, 07 Mar 2020 15:26:19 GMT	Sat, 07 Mar 2020 18:51:47 GMT	
ZTF20aasxxys	3	SN Hunter	Sat, 07 Mar 2020 15:32:54 GMT	Sat, 07 Mar 2020 18:49:56 GMT	
ZTF20aasyayz	3	SN Hunter	Sat, 07 Mar 2020 15:29:31 GMT	Sat, 07 Mar 2020 18:49:42 GMT	
ZTF20aasmawh	3	SN Hunter	Sat, 07 Mar 2020 15:23:46 GMT	Sat, 07 Mar 2020 18:49:15 GMT	
ZTF20aaszzez	3	SN Hunter	Sat, 07 Mar 2020 15:24:15 GMT	Sat, 07 Mar 2020 18:48:27 GMT	
ZTF20aatavub	3	SN Hunter	Sat, 07 Mar 2020 15:31:43 GMT	Sat, 07 Mar 2020 18:48:19 GMT	
ZTF20aatayji	3	SN Hunter	Sat, 07 Mar 2020 15:25:31 GMT	Sat, 07 Mar 2020 18:48:06 GMT	
ZTF20aatbznz	1	SN Hunter	Sat, 07 Mar 2020 18:36:01 GMT	Sat, 07 Mar 2020 18:36:01 GMT	
ZTF20aatckft	3	SN Hunter	Sat, 07 Mar 2020 15:24:53 GMT	Sat, 07 Mar 2020 18:35:00 GMT	

Figure 11. The ALeRCE reporter web interface (<http://reporter.alerce.online>) is used to manage user input in the ALeRCE ecosystem. Here we show two types of inputs: the reporter tool, which manages input labels from the SN hunter, either bogus (A) or possible SN (B), which in the latter case become candidates to be sent to the TOM; and the challenger tool, which we use to manage data classification challenges or hackathons. In the TOM list of possible SNe, users can select a given period of time of recently reported candidates (B1), which returns a given number of candidates (B2). Users can explore the object identifiers (B3), number of independent reports for the given candidate (B4), source of the label (B5), date of first (B6) and last (B7) reports, and possible actions (B8). Among the possible actions, users can explore who has reported a candidate (B9), create a target for observations in the TOM Toolkit (B10), edit the observational properties of an already-created TOM candidate (B11), or remove the target from the TOM Toolkit (B12). The full list can be downloaded as a CSV file (B13).

id	ra	dec
1	123.33041666666665	46.14694444444445
2	144.55124999999998	7.727777777777779
3	150.58374999999998	45.157500000000006
4	173.12125	3.958055555555557
5	194.81958333333333	55.251944444444455

Figure 12. Cross-match service interface (<https://xmatch.alerce.online>). Users can input arbitrary catalogs as CSV files to be cross-matched to the ZTF database. The procedure consists of selecting an input catalog CSV file (A) and then indicating the columns in the file that will be used as the identifier (B1), R.A. (B2), and decl. (B3), as well as the maximum radius used to search for the closest cross-matching source (B4). The information provided allows for the partial exploration of the input file (B5) by a given number of rows (B6) in paginated form (B7). After submitting the catalog (B8), users can visually explore and download the cross-matched catalog (C).

Python API Client. We provide a Python client for easier access to the previous API services. It can be installed via `pip` and is documented at <https://alerce.readthedocs.io/en/latest/>. You can find examples of how to use the client in the [use-case notebooks](#).

6. Results

The ALeRCE broker has processed 1.5×10^8 alerts from the public ZTF stream at a rate of about $5 \times 10^7 \text{ yr}^{-1}$, which corresponds to about $1.4 \times 10^5 \text{ night}^{-1}$, or about five alerts s^{-1} ,

on average. This is ~ 80 times less than the expected alert rate of LSST of about 10^7 night^{-1} . However, the ZTF public stream alert production rate is not constant, with some nights producing a few million alerts, which we have been able to ingest without significant wait-time increases. In Figures 14 and 15, we show the distribution of execution (CPU + input/output) and elapsed (including queue times and previous steps) times at the different steps of our pipeline for a typical ZTF night, including the distribution of ZTF streaming times (time between observation and ingestion) for comparison. With our

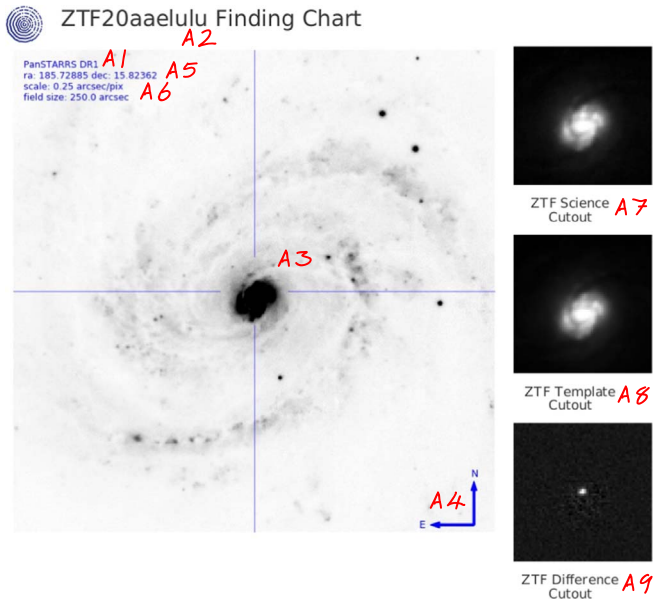


Figure 13. Section of the finding chart generated automatically for object ZTF20aaelulu, or SN 2020oi, a Type Ic SN that occurred in the nearby galaxy M100. The finding chart shows a PanSTARRS DR1 image (A1) centered around this object (A2, A3), indicating the direction of the north and east axes (A4), the coordinates (A5), and the pixel scale and field size (A6). It also shows the ZTF science (A7), reference (A8), and difference image stamps (A9). Additional information, such as the coordinates in a different format, magnitude statistics, or time of first and last detection, are also included. Note that this SN was reported to TNS by ALeRCE after being classified as a possible SN with just a single detection using the SN Hunter tool (see Figure 10).

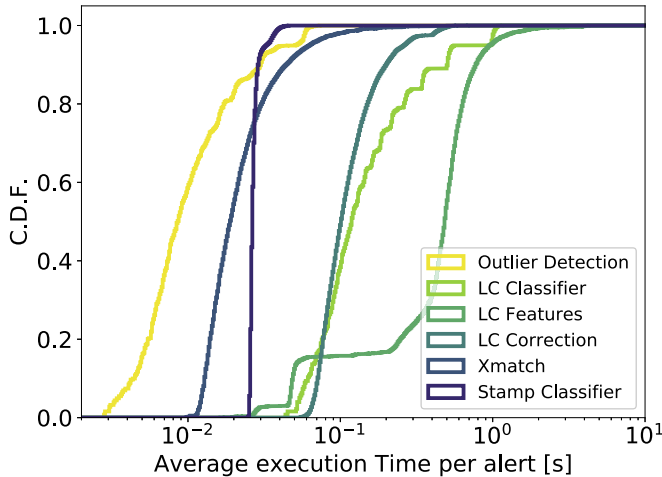


Figure 14. Cumulative distribution function of ALeRCE pipeline step average execution times, or the average time needed for an alert to be processed in a given step in batches, including only CPU and input/output times. In this figure, we consider an incoming alert rate of about 25 s^{-1} (we expect about 5 and 350 s^{-1} for ZTF and LSST, on average, respectively).

current infrastructure, we can process ZTF alerts in real time, with classification delays being dominated by the ZTF streaming times. The latest version of the ALeRCE pipeline has been tested at rates of about $400 \text{ alerts s}^{-1}$, which is more than the expected rate of LSST.

As of 2021 February, we had 5.1×10^7 objects, 1.5×10^8 detections, and 1.7×10^9 nondetections in our database. There are 1.1×10^6 objects classified by the light-curve classifier and 3.4×10^7 objects classified by the stamp classifier, which started

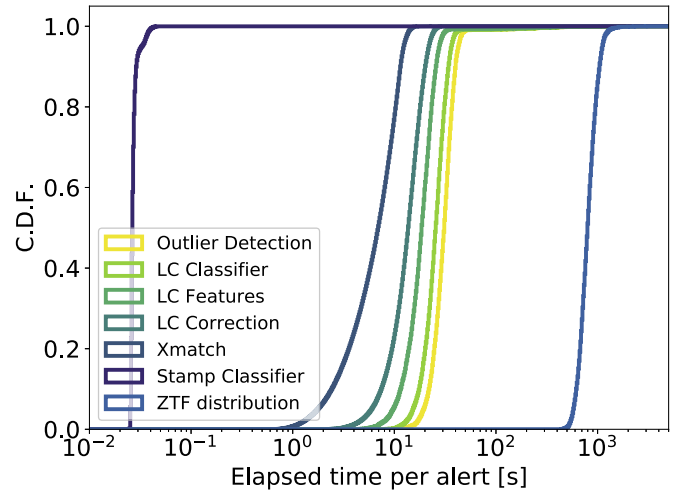


Figure 15. Same as Figure 14 but showing the cumulative distribution function of ZTF streaming times in comparison to the cumulative distribution function of ALeRCE pipeline elapsed times. The ZTF streaming times correspond to the difference between the reported observation time and the alert ingestion time, obtained empirically in a typical night of operations. The ALeRCE pipeline step elapsed time stands for the time needed for an alert to move from ingestion to the completion of a given step, including queue times. The difference between execution and average execution times is due mostly to the fact that some steps work with batches of alerts, increasing the efficiency, but also the queue times per alert. Note that in this experiment, we perform cross-matches after the stamp classifier.

being applied to new alerts in 2019 August. For a distribution of the ML inferred classes in these samples, see our accompanying papers (Carrasco-Davis et al. 2020; Sánchez-Sáez et al. 2021). The associated confusion matrices can be seen in Figure 3 and 4, and a comparison between the two classifiers can be seen in Figure 5. Note that our classifiers are continuously improving and that the choice of model is not based solely on a balanced accuracy score but also on a study of the relative frequency and spatial distribution of classes in the unlabeled set, which we have found to be an important verification when the training set is not representative of the unlabeled set.

An important tool to connect ALeRCE with the SN community of users is the SN Hunter. We have used it to report 6162 previously unreported astrophysical transient candidates to TNS, 883 of which have been classified spectroscopically (with less than 1% contamination among those classified spectroscopically; see Figure 16). Among these, we have found 128 SN candidates rising faster than 0.4 mag day^{-1} and 19 faster than 1.0 mag day^{-1} at discovery (see Figure 17). In the process, we have visually inspected 56,685 candidates, saving in our database 35,201 bogus candidates since 2019 October and 21,484 transient candidates since 2020 January, when we added the bogus and possible SN buttons to the SN Hunter, respectively. The bogus examples have been used to increase the size and diversity of our training set and resulted in significant improvements to the stamp classifier.

We are slowly building an international community of users. In order to facilitate the adoption of our tools by the community, we do not require users to create accounts to access our system, which makes it difficult to precisely estimate the number of ALeRCE users. However, we can use Google Analytics³⁵ to quantify our online community of users. Since 2019 July, when Google Analytics was added to the ALeRCE Explorer and SN Hunter tools, we have had 5066/1217 users

³⁵ <https://analytics.google.com>

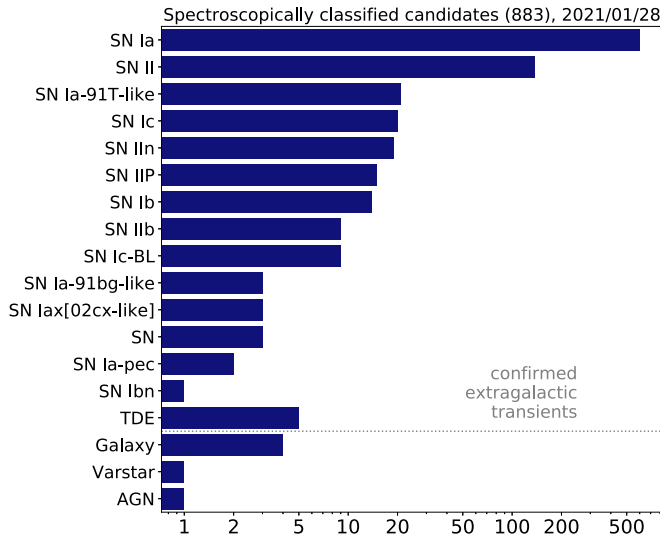


Figure 16. Sample of spectroscopically classified transients first reported by ALerCE to TNS from 6162 SN candidates submitted based on their first alert. Out of 883 candidates observed spectroscopically, 865 were classified as SNe, five as TDEs, two as unclear, four as galaxies (with one having an SN-like light curve), five as variable stars, one as an AGN, and one as other. Of the 865 confirmed SNe, 629 are SNe Ia, 171 are SNe II, 60 are SNe Iib/Ib/Ic, two are SLSNe, and three are classified as just SNe. The two unclear cases, both of which had SN-like light curves, are AT 2019yys (ZTF 19adcbnty), which could be an SN, TDE, or AGN, and AT 2020bdh (ZTF 20aaitvof), which has a very noisy spectrum.

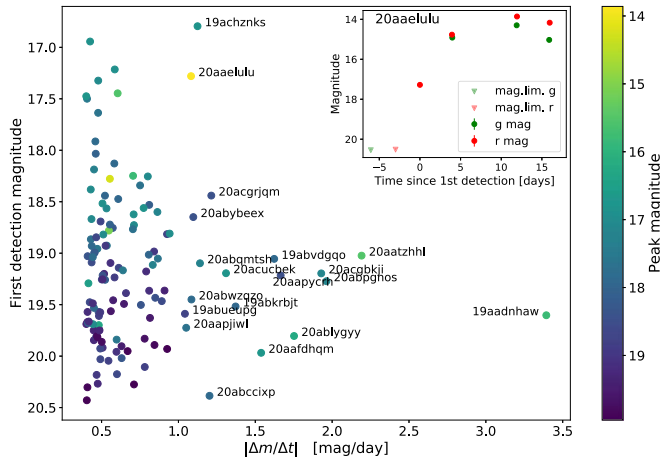


Figure 17. Detection magnitude vs. magnitude rise rate at time of detection for the SN candidates reported to TNS by ALerCE based on their first alert image stamps. The color indicates the peak magnitude of the candidate. We only show candidates detected rising faster than 0.4 mag day^{-1} , a sample that includes 128 SN candidates. We individually label 19 candidates that rose faster than 1 mag day^{-1} at detection. Of these candidates, ZTF 20abybeex, ZTF 20ablygyy, ZTF 20abccixp, ZTF 20aapycrh, ZTF 20aapjiwl, and ZTF19 abueupg are SNe II; ZTF 20aatzhhl and ZTF 20abwzqzo are SNe Iib; ZTF 19abvdgqo is an SN Ib; ZTF 20aaelulu is an SN Ic (shown in the inset plot); ZTF 20acubek, ZTF 20acgbkji, ZTF 20abqmtsh, and ZTF 19abkrbjt are SNe Ia; ZTF 19achznks and ZTF 20acgrjqm appear to be flaring AGNs; ZTF 20aafdhqm is a transient that coincided with a previous SN candidate (PS1-13dgc); and ZTF 19aadnhaw and ZTF 20abpgnos are probably novae based on their light curves.

(unique combinations of device and browser, as per the Google definition) and 23,672/6218 sessions in the ALerCE Explorer/SN Hunter. This does not include the use of APIs or direct connections to our database. Our users are currently distributed in 66 countries (see Figure 18), with the top ones being the U.S.

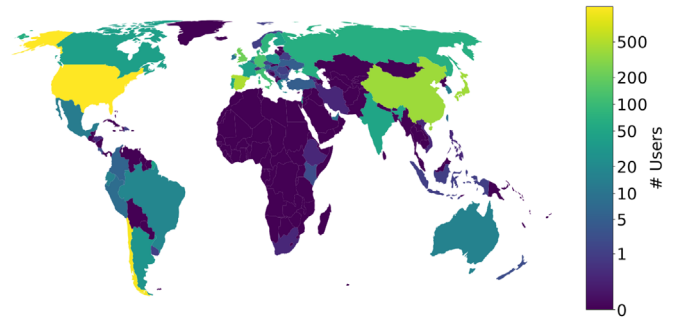


Figure 18. Geographic distribution of users of the ALerCE Explorer according to Google Analytics. The number of users is estimated by counting the unique combinations of devices and browsers accessing our website. In total, there are 5066 estimated users coming from 66 different countries accessing the ALerCE Explorer.

(25.4%), Chile (21.9%), Spain (9.9%), China (8.2%), Japan (8.0%), and the U.K. (4.7%). We are continuously listening to our users via tutorial workshops to include new features and create new use-case Jupyter notebooks for different science cases. We encourage users to create additional use-case notebooks and contribute to our open-source repository (<https://github.com/alercebroker/usecases>).

7. Discussion and Conclusions

The ALerCE broker is a new-generation astronomical alert broker processing alerts in real time from ZTF and preparing to become a community broker for LSST. We are an interdisciplinary, interinstitutional, and international team led from Chile using Agile methodologies to develop new digital components for the astronomical time-domain ecosystem in the era of large extendable telescopes.

In this paper, we have reported the motivation, challenges, methodologies, and first results of the ALerCE broker. The main motivation for ALerCE is to provide a rapid classification of events to enable fast follow-up and characterization, but it is also intended to provide a systematic classification of all variable objects for a self-consistent analysis of large volumes of events in the observable universe. Our primary scientific drivers are the study of transients, variable stars, and AGNs, but we also provide solar system object classifications for further analysis.

We describe the infrastructure, processing steps, data products, tools, and services that work in real time. We ingest, aggregate, and cross-match the alert stream and apply two ML-based classifiers to the data (see Section 3). First, a stamp classifier is applied to all alerts associated with previously unreported objects using the first image stamps as input and a simple taxonomy. Second, a light-curve classifier with a more complex taxonomy is applied to all objects with ≥ 6 detections in g or r. We are also experimentally applying outlier detection methods to the data, which we hope to make public in real time after significant testing is done. To our knowledge, ALerCE was the first public broker to provide real-time classification of the ZTF alert stream into an astrophysically motivated taxonomy based on the alert image stamps or their light curves.

Regarding the processing of the data, our processing times per alert are of the order of tens of seconds, significantly smaller than the current ZTF streaming times (see Section 6). Moreover, we have run experiments at ingestion rates similar to those expected for LSST.

Our database contains object-, detection-, and nondetection-based families of tables with increasing numbers of rows that are indexed for fast query speeds. All relevant tables are public with read-only access, although we recommend accessing them via our different APIs that power all of our web-based services and Python client. We provide extensive documentation for our different data products and services, which can be found in our main website, <http://alerce.science>. All of our data products, documentation, tools, and services are summarized in Table 3.

Apart from providing a classified stream of data upon request, our two most important web services are the ALerCE Explorer (<https://alerce.online>) and the SN Hunter (<https://snhunter.alerce.online>), which are publicly available and described in detail in Section 5.2.1. The ALerCE Explorer is the main tool to explore the objects contained in the ZTF public stream, allowing for simple queries and providing a user-friendly visualization of their light curves, cross-matches, image stamps, and classification probabilities. The SN Hunter tool is targeted for the transient community to enable a rapid reaction, allowing users to quickly explore and provide feedback on the latest SN candidates contained in the stream. We use this tool to submit new SN candidates to the TNS at an average rate of about nine night⁻¹, with 6162 reported candidates since 2019 August. We also use this tool to select candidates for follow-up via the TOM Toolkit.

An important goal of ALerCE is to provide a good user experience, which should allow for a smooth transition into a time-domain ecosystem dominated by large alert streams and automated components where astronomers and data scientists are not replaced but instead aided by ML tools to achieve new discoveries. Thus, we are developing different modular components for the visualization of the alert stream data, optimized for usability after testing with our community of users in regular tutorials and hackathons. The use of Agile methodologies with a fully dedicated interdisciplinary team of engineers and astronomers has been critical to develop ALerCE at the speed required by the community. Collaboration remains essential among brokers to bring a more diverse set of ideas into our community and add resilience to the time-domain ecosystem in the era of large etendue telescopes.

One of the biggest challenges ahead for ALerCE is the ability to scale to significantly larger streams, from $\sim 1.4 \times 10^5$ to $>10^7$ alerts night⁻¹, and with significantly more objects generating alerts, from a few 10^7 to $>10^9$ objects. For this, we will migrate some of our tables from an SQL centralized database engine to a NoSQL distributed database engine (e.g., Cassandra, MongoDB). We are running different tests to determine the efficiency and cost of the different available solutions in collaboration with other brokers (Fink). We have performed experiments at rates larger than expected for LSST for the messaging system (4000 messages s⁻¹), processing steps (400 messages s⁻¹), database insertions (1.8×10^5 messages s⁻¹), and database spatial queries (10^4 messages s⁻¹). Another important challenge is to determine what fraction of our storage and computing services should be located in the cloud (e.g., AWS, where we currently operate some of our services) versus on-premise infrastructure. It seems likely that the answer will be a hybrid solution, with cloud and on-premise infrastructure optimized for a better user experience while minimizing the operational costs.

Achieving more complex taxonomies in an era of multi-stream, multimessenger astronomy is another important

challenge ahead. In fact, the large number of events expected, combined with the addition of heterogeneous streams spanning different depths, cadences, wavelengths, and messengers, will likely unveil new populations that would not have been possible to identify otherwise. Encompassing the full diversity of variable classes in the universe with a fixed taxonomy is unfeasible; thus, our taxonomy will continue to grow and evolve with time. Eventually, a combination of domain knowledge, via supervised training, and unsupervised, more data-driven taxonomies will become necessary. Training and classifying with missing data, as most streams of data will be sparse in comparison to that of LSST, will also become important.

Regarding the challenges of ML classification, we are trying different strategies. We are introducing new features, e.g., a complex number extension to the IAR model that allows for positive as well as negative autocorrelation (Elorrieta et al. 2019) further expanded to bivariate or higher-dimensional time series and includes different covariance structures. From these models, we expect to extract useful features for classification, as well as be able to do prediction, interpolation, and forecasting on time series. We are also testing methods to combine real, augmented, and simulated data; new ways to combine and expand our stamp and light-curve classifiers; or different recurrent neural networks applied to the light curve (e.g., Muthukrishna et al. 2019) and image stamp series (e.g., Carrasco-Davis et al. 2019); or different outlier detection methods.

Finally, we note that, given the continuously evolving nature of ALerCE, this paper provides a snapshot of the current status of ALerCE as of 2021 February. We are constantly listening to our community of users in an effort to introduce new data products, tools, and services. Some of the services under development include custom-made filters for the alert stream via extension of the SN Hunter, watch lists, batch processing tools, and a simple tool to perform complex queries in our database. Our preferred way of communication is through issues in our GitHub repositories (<https://www.github.com/alercebroker>), but users can also contact us directly via <https://alerce.science>.

This work was funded by ANID—Millennium Science Initiative Program—ICN12_009 awarded to the Millennium Institute of Astrophysics MAS (A.C., A.M., A.M.A., A.S., C.S.C., C.D.O., C.V., D.D.C., D.R.Ma., D.R.Mi., E.C.N., E.R., F.E., F.E.B., F.F., G.C.V., G.P., I.A.M., I.R., J.A., J.B., J.R.V., L.H.G., L.S.G., M.C., M.P.C., N.A., P.A.E., P.H., P.S.S., R.C.D., S.E., R.K., and W.P.), and National Agency for Research and Development (ANID) grants: Basal Center for Mathematical Modeling grant CMM ANID PIA AFB170001 (A.M., A.M.A., C.V., C.S.C., E.C.N., E.V., D.R.Ma., D.R.Mi., F.F., I.A.M., I.R., J.C.M., J.S.M., L.S.G., and P.A.E.); Centro de Astrofísica y Tecnologías Afines AFB-170002 (D.D.C., F.E.B., M.C., P.S.S., and A.C.); FONDECYT Regular Nos. 1200710 (F.F.), 1190818 (F.E.B.), 1200495 (F.E.B.), 1171273 (M.C.), 1201793 (G.P.), and 1171678 (P.A.E.); FONDECYT Iniciación Nos. 11200590 (F.E.) and 11191130 (G.C.V.); FONDECYT Postdoctorado Nos. 3200250 (P.S.S.) and 3200222 (D.D.C.); Magister Nacional 2019 No. 22190947 (E.R.); and ANID infrastructure funds QUIMAL140003 and QUIMAL190012. We acknowledge support from REUNA Chile, which hosts and

maintains some of our infrastructure. This work has been possible thanks to the use of AWS-U.Chile-NLHPC credits. This work was funded in part by project CORFO 10CEII-9157 Inria Chile. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). This project was supported by the Competition for Research Regular Projects, year 2019, code LPR19-22, Universidad Tecnológica Metropolitana and the high-performance computing system of PIDi-UTEM (SCC-PIDi-UTEM—CONICYT—FONDEQUIP—EQM180180).

Software: Aladin (Bonnarel et al. 2000), Apache ECharts,³⁶ Apache Kafka,³⁷ Apache Spark (Zaharia et al. 2016), ASTROIDE (Brahem et al. 2018), Astropy (Astropy Collaboration et al. 2013), catsHTM (Soumagnac & Ofek 2018), Dask (Rocklin 2015), FATS (Nun et al. 2017), Grafana,³⁸ Imbalanced-learn (Lemaître et al. 2017), ipyladin (Boch & Desroziers 2020), Jupyter (Kluyver et al. 2016), Keras (Gulli et al. 2017), Matplotlib (Hunter 2007), NED (Steer et al. 2017), P4J (Huijse et al. 2018), Pandas (McKinney et al. 2010), Prometheus,³⁹ Python (Van Rossum & Drake 1995), scikit-learn (Pedregosa et al. 2011), Simbad-CDS (Wenger et al. 2000), Tensorflow (Abadi et al. 2016), Vue,⁴⁰ Vuetify,⁴¹ PostgreSQL,⁴² XGBoost.⁴³

Appendix A Light-curve Correction Derivation

A.1. Light-curve Fluxes

An alert is originated when a significant flux is detected at some location of a difference image between a science and a reference image. In the ZTF alert stream, the difference and reference fluxes are reported for every alert. The science flux is not reported, but it can be recovered from the difference and reference images. The difference flux is reported by its absolute magnitude, m_{diff} , and sign, sgn , and the reference flux is reported by the PSF photometry magnitude, m_{ref} , of the closest source in the reference with associated errors, distance, and shape parameters. This leads to three types of cases: (1) the closest source in the reference coincides with the location of the alert, and it is unresolved; (2) the closest source in the reference coincides with the location of the difference image alert, but it is resolved; and (3) the closest source does not coincide with the position of the difference alert. In case (1), the science flux can be recovered exactly; in case (2), it can be recovered plus a constant that depends on how much contamination from an extended source occurs in the reference; and in case (3), one needs to assume that the science flux is equal to the difference flux. These cases are typically represented by variable stars (1), AGNs (2), or transients (3). Since it is not possible to know a priori which correction should be applied to each object, e.g., it is difficult to distinguish an AGN from a nuclear transient until the flux evolution can be observed, we report both the corrected photometry, which is useful for variable stars and AGNs, and the uncorrected photometry, which is useful for transients.

If the reference source is resolved, its reported flux contains two components: a variable/compact component, which is normally the object of study, and a static/extended component, which is difficult to separate using only the ZTF photometry. Because of the convolution done during the image difference process, the extended component should not contribute to the difference flux. Then, we note the following relations:

$$f_{\text{ref}} = f_{\text{ref}}^{\text{ext}} + f_{\text{ref}}^{\text{var}}, \quad (\text{A1})$$

$$f_{\text{sci}} = f_{\text{sci}}^{\text{ext}} + f_{\text{sci}}^{\text{var}}, \quad (\text{A2})$$

$$\text{sgn} f_{\text{diff}} = f_{\text{sci}}^{\text{var}} - f_{\text{ref}}^{\text{var}}, \quad (\text{A3})$$

where f_{ref} is the reference flux, f_{sci} is the science flux, sgn is the sign, f_{diff} is the absolute value of the difference flux, $f_{\text{ref}}^{\text{ext}}$ is the contribution from the extended component in the reference image, $f_{\text{ref}}^{\text{var}}$ is the contribution of the variable component in the reference image, $f_{\text{sci}}^{\text{ext}}$ is the contribution from the extended component in the science image, and $f_{\text{sci}}^{\text{var}}$ is the contribution of the variable component in the science image. Note that the contribution of the extended component can vary between the reference and science images due to seeing effects, which can create an artificial source of variability. The scientifically relevant component for variability studies is the flux of the compact component, but it is difficult to separate it from the extended component. The second-best alternative is to recover the flux of the compact component plus a constant contribution from the extended component. For this, we can define an effective science flux, \hat{f}_{sci} ,

$$\hat{f}_{\text{sci}} \equiv f_{\text{ref}}^{\text{ext}} + f_{\text{sci}}^{\text{var}}, \quad (\text{A4})$$

$$= f_{\text{ref}} + \text{sgn} f_{\text{diff}}, \quad (\text{A5})$$

which considers the same contribution of the extended component at all times. If the reference image changes, we can introduce a new effective science flux, $\hat{f}_{\text{ref}, 0}$, that considers the contribution from the extended component from the first reference image used to generate alerts,

$$\hat{f}_{\text{sci}, 0} = f_{\text{ref}, 0}^{\text{ext}} + f_{\text{sci}}^{\text{var}}, \quad (\text{A6})$$

$$= \hat{f}_{\text{sci}} + (f_{\text{ref}, 0}^{\text{ext}} - f_{\text{ref}}^{\text{ext}}), \quad (\text{A7})$$

where $f_{\text{ref}, 0}^{\text{ext}}$ is the (unknown) contribution from the extended component from the first reference image. Note that the expected value from the second term is zero.

A.2. Light-curve Variances

The computation of the errors of the science flux must take into account the relation between the difference and reference fluxes, which are correlated. We can estimate the variance of the effective science flux, $\mathbb{V}[\hat{f}_{\text{sci}}]$, starting from Equation (A5) and using Equations (A1) and (A3):

$$\mathbb{V}[\hat{f}_{\text{sci}}] = \mathbb{V}[f_{\text{ref}} + \text{sign} f_{\text{diff}}] \quad (\text{A8})$$

$$= \mathbb{V}[f_{\text{ref}}] + \mathbb{V}[f_{\text{diff}}] + 2 \text{Cov}[f_{\text{ref}}, \text{sign} f_{\text{diff}}] \quad (\text{A9})$$

$$= \mathbb{V}[f_{\text{ref}}] + \mathbb{V}[f_{\text{diff}}] + 2 \text{Cov}[f_{\text{ref}}^{\text{ext}} + f_{\text{ref}}^{\text{var}}, f_{\text{sci}}^{\text{var}} - f_{\text{ref}}^{\text{var}}] \quad (\text{A10})$$

$$= \mathbb{V}[f_{\text{ref}}] + \mathbb{V}[f_{\text{diff}}] - 2 \mathbb{V}[f_{\text{ref}}^{\text{var}}]. \quad (\text{A11})$$

³⁶ <https://echarts.apache.org>

³⁷ <https://kafka.apache.org/>

³⁸ <https://grafana.com/>

³⁹ <https://prometheus.io/>

⁴⁰ <https://vuejs.org/>

⁴¹ <https://vuetifyjs.com/>

⁴² <https://www.postgresql.org/>

⁴³ <https://xgboost.readthedocs.io/>

Note that the variance due to sky emission is contained in the first two terms of Equation (A16). One can also include additional terms in Equation (A10) to reflect the contribution of the sky, but because these terms are not correlated, they have no additional contribution in the covariance. We can expand Equation (A11) to get the following:

$$\mathbb{V}[\hat{f}_{\text{sci}}] = \mathbb{V}[f_{\text{ref}}] + \mathbb{V}[f_{\text{diff}}] - 2 \mathbb{V}[f_{\text{ref}}^{\text{var}}] \quad (\text{A12})$$

$$= \mathbb{V}[f_{\text{ref}}^{\text{ext}} + f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{diff}}] - 2 \mathbb{V}[f_{\text{ref}}^{\text{var}}] \quad (\text{A13})$$

$$= \mathbb{V}[f_{\text{ref}}^{\text{ext}}] + \mathbb{V}[f_{\text{ref}}^{\text{var}}] + 2 \text{Cov}[f_{\text{ref}}^{\text{ext}}, f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{diff}}] - 2 \mathbb{V}[f_{\text{ref}}^{\text{var}}] \quad (\text{A14})$$

$$= \mathbb{V}[f_{\text{ref}}^{\text{ext}}] + \mathbb{V}[f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{diff}}] - 2 \mathbb{V}[f_{\text{ref}}^{\text{var}}] \quad (\text{A15})$$

$$= \mathbb{V}[f_{\text{diff}}] - \mathbb{V}[f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{ref}}^{\text{ext}}], \quad (\text{A16})$$

and, in the case of a change in the reference image, using Equations (A7), (A16), and (A4):

$$\mathbb{V}[\hat{f}_{\text{sci}, 0}] = \mathbb{V}[\hat{f}_{\text{sci}} + (f_{\text{ref}, 0}^{\text{ext}} - f_{\text{ref}}^{\text{ext}})] \quad (\text{A17})$$

$$= \mathbb{V}[\hat{f}_{\text{sci}}] + \mathbb{V}[f_{\text{ref}, 0}^{\text{ext}}] + \mathbb{V}[f_{\text{ref}}^{\text{ext}}] - 2 \text{Cov}[\hat{f}_{\text{sci}}, f_{\text{ref}}^{\text{ext}}] \quad (\text{A18})$$

$$= \mathbb{V}[f_{\text{diff}}] - \mathbb{V}[f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{ref}}^{\text{ext}}] + \mathbb{V}[f_{\text{ref}, 0}^{\text{ext}}] + \mathbb{V}[f_{\text{ref}}^{\text{ext}}] - 2 \text{Cov}[f_{\text{ref}}^{\text{ext}} + f_{\text{ref}, 0}^{\text{ext}}, f_{\text{ref}}^{\text{ext}}] \quad (\text{A19})$$

$$= \mathbb{V}[f_{\text{diff}}] - \mathbb{V}[f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{ref}, 0}^{\text{ext}}]. \quad (\text{A20})$$

To summarize, we show Equations (A5), (A7), (A16), and (A20):

$$\begin{aligned} \hat{f}_{\text{sci}} &= f_{\text{ref}} + \text{sgn} f_{\text{diff}} \\ \hat{f}_{\text{sci}, 0} &= \hat{f}_{\text{sci}} + (f_{\text{ref}, 0}^{\text{ext}} - f_{\text{ref}}^{\text{ext}}) \\ \mathbb{V}[\hat{f}_{\text{sci}}] &= \mathbb{V}[f_{\text{diff}}] - \mathbb{V}[f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{ref}}^{\text{ext}}] \\ \mathbb{V}[\hat{f}_{\text{sci}, 0}] &= \mathbb{V}[f_{\text{diff}}] - \mathbb{V}[f_{\text{ref}}^{\text{var}}] + \mathbb{V}[f_{\text{ref}, 0}^{\text{ext}}]. \end{aligned}$$

A problem with these formulae is that neither the variable nor the extended components are known. However, they led us to consider the following cases.

1. The contribution from the extended component is negligible in all reference images:

$$f_{\text{ref}}^{\text{ext}} = 0 \Rightarrow$$

$$\hat{f}_{\text{sci}, 0} = \hat{f}_{\text{sci}} = f_{\text{ref}} + \text{sgn} f_{\text{diff}}, \quad (\text{A21})$$

$$\mathbb{V}[\hat{f}_{\text{sci}, 0}] = \mathbb{V}[\hat{f}_{\text{sci}}] = \mathbb{V}[f_{\text{diff}}] - \mathbb{V}[f_{\text{ref}}]. \quad (\text{A22})$$

2. The contribution from the extended component is similar in all reference images, and its contribution is similar to that from the variable component:

$$f_{\text{ref}, 0}^{\text{ext}} = f_{\text{ref}}^{\text{ext}} \& f_{\text{ref}}^{\text{var}} = f_{\text{ref}}^{\text{ext}} \Rightarrow$$

$$\hat{f}_{\text{sci}, 0} = \hat{f}_{\text{sci}} = f_{\text{ref}} + \text{sgn} f_{\text{diff}}, \quad (\text{A23})$$

$$\mathbb{V}[\hat{f}_{\text{sci}, 0}] = \mathbb{V}[\hat{f}_{\text{sci}}] = \mathbb{V}[f_{\text{diff}}]. \quad (\text{A24})$$

3. The contribution from the extended component is similar in all reference images, and its contribution is dominant

over the variable component:

$$f_{\text{ref}, 0}^{\text{ext}} = f_{\text{ref}}^{\text{ext}} \& f_{\text{ref}}^{\text{var}} = 0 \Rightarrow$$

$$\hat{f}_{\text{sci}, 0} = \hat{f}_{\text{sci}} = f_{\text{ref}} + \text{sgn} f_{\text{diff}}, \quad (\text{A25})$$

$$\mathbb{V}[\hat{f}_{\text{sci}, 0}] = \mathbb{V}[\hat{f}_{\text{sci}}] = \mathbb{V}[f_{\text{diff}}] + \mathbb{V}[f_{\text{ref}}]. \quad (\text{A26})$$

A visual inspection of variable starlight curves confirms that Equation (A22) is a better approximation in the case where there is no contribution from an extended component. In the case of AGNs, we have found that Equation (A24) appears to be a better reflection of the measurement errors, which is consistent with having a similar contribution from the extended and variable components. In the case of transients, the extended component dominates the flux in the reference, but for these cases, the scientifically relevant flux is the difference flux and its error. For this reason, we report the difference flux with its error, as well as the effective science flux with the errors (after a conversion of the fluxes to magnitudes) from Equations (A22) and (A24) for every object where it is possible to correct the photometry, letting the users decide which flux and error to use for their particular science.

A.3. Light-curve Magnitudes

The corrected photometry magnitude results from adding/ subtracting the fluxes from the reference and difference in the same unit system and then converting to magnitudes. We can compute \hat{f}_{sci} by transforming the reference and difference magnitudes using the zero-points of the science image,

$$\begin{aligned} \hat{f}_{\text{sci}} &= f_{\text{ref}} + \text{sgn} f_{\text{diff}} = 10^{\frac{\text{ZP}_{\text{sci}} - m_{\text{ref}}}{2.5}} \\ &\quad + \text{sgn} 10^{\frac{\text{ZP}_{\text{sci}} - m_{\text{diff}}}{2.5}}, \end{aligned} \quad (\text{A27})$$

where ZP_{sci} is the zero-point of the science image. This implies that the effective science magnitude, \hat{m}_{sci} , will be

$$\begin{aligned} \hat{m}_{\text{sci}} &= -2.5 \log f_{\text{sci}} + \text{ZP}_{\text{sci}} \\ &= -2.5 \log (10^{\frac{\text{ZP}_{\text{sci}} - m_{\text{ref}}}{2.5}} + \text{sgn} 10^{\frac{\text{ZP}_{\text{sci}} - m_{\text{diff}}}{2.5}}) + \text{ZP}_{\text{sci}} \\ &= -2.5 \log (10^{-\frac{m_{\text{ref}}}{2.5}} + \text{sgn} 10^{-\frac{m_{\text{diff}}}{2.5}}). \end{aligned} \quad (\text{A28})$$

Finally, we show the reported errors for Equations (A22) and (A24),

$$\delta \hat{m}_{\text{sci}} = \frac{(10^{-0.8 m_{\text{diff}}} \delta m_{\text{diff}}^2 - 10^{-0.8 m_{\text{ref}}} \delta m_{\text{ref}}^2)^{0.5}}{10^{-0.4 m_{\text{ref}}} + \text{sgn} 10^{-0.4 m_{\text{diff}}}}, \quad (\text{A29})$$

to be used when there is no significant contribution from an extended component, or

$$\delta \hat{m}_{\text{sci}} = \frac{10^{-0.4 m_{\text{diff}}} \delta m_{\text{diff}}}{10^{-0.4 m_{\text{ref}}} + \text{sgn} 10^{-0.4 m_{\text{diff}}}}, \quad (\text{A30})$$

to be used when there is a contribution from an extended component assumed to be similar to the variable component.

Table 4 provides the list of telescopes that were used in preparing Figure 1, along with their names and a relevant accompanying reference.

Tables 5, 6, and 7 refer to a number of studies in which light curves were used to perform ML-based classification of variable and transient sources. Tables 5 and 6 both refer to studies in which only persistent variable star classes were used;

Table 4
Selection of Telescopes Shown in Figure 1

Short Name	Long Name	Reference
ASAS-SN	All-Sky Automated Survey for Supernova	Kochanek et al. (2017)
ATLAS	Asteroid Terrestrial-impact Last Alert System	Tonry et al. (2018)
BlackGEM	BlackGEM	https://astro.ru.nl/blackgem/
Blanco-DECam	Víctor Blanco telescope—Dark Energy Camera	Flaugher et al. (2015)
Clay-MegaCam	Clay Telescope—Megacam	McLeod et al. (2015)
CFHT-MegaCam	Canada–France–Hawaii Telescope—Megacam	Boulade et al. (2003)
CRTS	Catalina Real-Time Transient Survey (CSS, MLS, SSS)	Drake et al. (2009)
Euclid	Euclid Mission	Laureijs et al. (2011)
Evryscope	Evryscope—South	Law et al. (2015)
Gaia	Gaia mission	Gaia Collaboration et al. (2018)
HATPI	HATPI	https://hatpi.org/science/
Kepler	Kepler mission	Borucki et al. (2010)
KMTNet	Korea Microlensing Transient Network	Kim et al. (2016)
KISO	Kiso Observatory	Morokuma et al. (2014)
LS-QUEST	La Silla 40" ESO Schmidt telescope—QUEST camera	Vivas et al. (2004)
LSST	Vera C. Rubin Observatory Legacy Survey of Space and Time	LSST Science Collaboration et al. (2009)
PanSTARRS	Panoramic Survey Telescope and Rapid Response System	Kaiser et al. (2002)
PTF	Palomar Transient Factory	Law et al. (2009)
SDSS	Sloan Digital Sky Survey	York et al. (2000)
Subaru-HSC	Subaru telescope—Hyper Suprime-Cam	Aihara et al. (2018)
SkyMapper	SkyMapper Southern Sky Survey	Keller et al. (2007)
TESS	Transiting Exoplanet Survey Satellite	Ricker et al. (2015)
VISTA	Visible and Infrared Survey Telescope for Astronomy	Dalton et al. (2006)
VST-OmegaCam	VLT Survey Telescope—OmegaCam	Cappellaro (2005)
WFIRST	Wide Field Infrared Survey Telescope (aka Nancy Grace Roman Space Telescope)	Spergel et al. (2015)
ZTF	Zwicky Transient Facility	Bellm et al. (2019)

Table 5
Light Curve–based ML Classifiers that Include Only Persistent Variable Classes (More than Two Classes) between 2017 and 2019

Reference	Data Source	No. of Classes	Classes
Rimoldini et al. (2019)	Gaia DR2	18	E, CV, RSCvn, BLAP, Mira+SR, DSCT+SXPh, RRL(ab, c, d, Ad), CephCl, ACEP, CephII, Low amp.:DSCT+GDOR, ELL, OSARG, FL+ROT, other
Tsang & Schultz (2019)	ASAS-SN	8	DSCT, RRL(ab, cd), Ceph, E, ROT, Mira, SR
Jayasinghe et al. (2019)	ASAS-SN	10	Ceph, DSCT, E(EW, EA EB, EB), RRL(ab,c), M, SR, Irregular
Hosenie et al. (2019)	CSDR2	12	RRL(ab, c, d), Blazhko, E(C+SD, D), ROT, LPV, DSCT, Ceph(II, A)
Johnston et al. (2020)	UCR	3	RRL, Ceph, E
	LINEAR	5	RRL(ab, c), DSCT, E(C, SD)
Aguirre et al. (2019)	OGLE+VVV +CoRoT	9	Ceph(F, O1), RRL(ab, c), E(C, SD+D), Mira, SR, OSARG
Castro et al. (2018)	MACHO	8	NV, QSO, BeS, Ceph, RRL, E, ML, LPV
	OGLE	6	Ceph, CephII, RRL, E, DSCT, LPV
Naul et al. (2018)	ASAS	5	RRLab, Ceph, SR, BPer, WUMa
	LINEAR	5	DSCT, RRL(ab, c), BPer, WUMa
	MACHO	8	Ceph(F, O1), LPVW, RRL(ab, c, e, GB)
Valenzuela & Pichara (2018)	OGLE	8	Ceph(CL, II, A), RRL, LPV, DPV, DSCT, E
	MACHO	11	RRL(ab, c, e, GB), Ceph(F, O1), LPVW(A, B, C, D), E
Mahabal et al. (2017)	CSDR2	7	E(C, SD), RRL(ab, c, d), RSCVn, LPV
Benavente et al. (2017)	EROS, MACHO, HiTS	5	Ceph, E, QSO, RRL, LPV
Zinn et al. (2017)	OGLE	8	Mira, QSO, SR, OSARG, Ceph(F, O1), RRL(ab+d, c+e)

Note. Class abbreviations are defined in Tables 10–14.

Table 6
Light Curve–based ML Classifiers that Include Only Persistent Variable Objects (More than Two Classes) before 2017

Reference	Data Source	No. of Classes	Classes
Kim & Bailer-Jones (2016)	MACHO, LINEAR, ASAS	19	DSCT, RRL(ab, c, d, e), Ceph(F, O1, other, II), E(C, SD, D), LPV(MAGBC, MAGBO, OSARGAGB, OSARGRGB, SRAGBC, SRAGBO), NV
Mackenzie et al. (2016)	OGLE	6	Ceph(CL, II), RRL, E, DSCT, LPV
Pichara et al. (2016)	MACHO	8	NV, QSO, BeS, Ceph, RRL, E, ML, LPV
	MACHO	8	BeS, Ceph, E, LPV, ML, NV, QSO, RRL
	EROS	11	E, RRL, Ceph(F, O1, DM, II), LPV(OSARGRGBO, SRAGBO, SRAGBC, MAGBC, MAGBO)
Nun et al. (2016)	MACHO	8	NV, QSO, BeS, Ceph, RRL, E, ML, LPV
Bass & Borne (2016)	Kepler	14	ACT, BCep, Ceph, DSCT, E, ELL, GDor, ROT, RRL(ab, c), RVTau, SPB, SR, MISC/NV
Faraway et al. (2016)			
Kügler et al. (2015)	OGLE	3	Ceph, E, RRL
	ASAS	7	Mira, RRLab, E(C, D, SD), DSCT, CephF
Kim et al. (2014)	EROS-2	26	DSCT, RRL(ab, c, d, e), Ceph(F, O1, other), CephII E(C, SD, D, SD+D, other), BeS, QSO, NV LPV(MAGB(C, O), OSARGAGB(C, O), OSARGRGB(C, O), SRAGB(C, O))
Pichara & Protopapas (2013)	SAGE, 2MASS, UBVI, MACHO	7	NV, QSO, BeS, Ceph, RRL, E, LPV
Richards et al. (2012)	ASAS	28	DSCT, SXPh, RRL(ab, c, d), Ceph(CL, MM, II), Mira, SR, LPVW(A, B), RVTau, BCep, RSG, BPer, BLyr, WUMa, ChemPec, ELL, RSCvn, HAeBe, CTTau, WLTTau, RCB, LBV, BeS
Debosscher et al. (2009)	CoRoT	29	sdBV, DSCT, LBoo, SXPh, roAp, GDor, RR(ab, c, d), Ceph(CL, DM, II), RVTau, Mira, SR, PVSG, BCep, SPB, E, ChemPec, ELL, FUOri, HAeBe, TTau, LBV, WR, XB, BeS, LAPV
Debosscher et al. (2007)	OGLE	35	DAV, DBV, sdBV, GWVir, DSCT, LBoo, SXPh, roAp, GDor, RRL(ab, c, d), Ceph(CL, DM, II), PVSG, Mira, SR, RVTau, BCep, SPB, E(C, SD, D), ChemPec, ELL, FUOri, HAeBe, TTau, LBV, SLR, WR, XB, CV, BeS

Note. Class abbreviations are defined in Tables 10–14.

Table 7
Light Curve–based ML Classifiers that Include Only Transient Objects

Reference	Data Source	No. of Classes	Classes
Villar et al. (2019)	PS1-MDS	5	SN Ia, SN Ibc, SN II, SN IIIn, SLSN
Muthukrishna et al. (2019)	PLAsTiCC	12	TDE, CART, ILOT, PISN, kN, .Ia, SN Ia, SN Iax, SN Ia-91bg, SN Ibc, SN II
Möller & de Boissière (2020)	SNANA	2	SN Ia, other
Brunel et al. (2019)	SNANA, SPCC	2	SN Ia, other
Revsbech et al. (2018)	SPCC	3	SN Ia, SN II, SN Ibc
Charnock & Moss (2017)	SPCC	3	SN Ia, SN II, SN Ibc
Lochner et al. (2016)	SPCC	3	SN Ia, SN II, SN Ibc
Karpenka et al. (2013)	SPCC	2	SN Ia, other

Note. Class abbreviations are defined in Table 14.

the former refers to papers published between 2017 and 2019, whereas the latter includes studies that appeared in print before 2017. Table 7, in turn, refers to those studies in which only transient sources were considered. These three tables have the same structure, with the reference given in the first column, an acronym for the source of the data given in the second column

(with keys provided in Tables 8 and 9 for empirical and synthetic data, respectively), the number of classes considered shown in the third column, and the fourth column displaying acronyms representing the actual classes that were considered in each case. These acronyms, along with the classes that they are intended to represent, are laid out in Tables 10–14.

Table 8
Observational Data Sources Used for ML Classification

Abbreviation	Long Name	Reference
ZTF	Zwicky Transient Facility	Bellm et al. (2019)
HSC-SSP	Hyper Suprime-Cam Subaru Strategic Program	Aihara et al. (2018)
UCR	University of California Riverside Time Series Classification Archive	Dau et al. (2018)
OSC	Open Supernova Catalog	Guillochon et al. (2017)
ASAS-SN	All-Sky Automated Survey for Supernovae	Kochanek et al. (2017)
CSDR2	Catalina Surveys Data Release 2	Drake et al. (2017)
HiTS	High cadence Transient Survey	Förster et al. (2016)
PS1-MDS	PanSTARRS-1 Medium Deep Survey	Huber et al. (2011)
LINEAR	Lincoln Near-Earth Asteroid Research Survey	Sesar et al. (2011)
UBVI	UBVI photometry of six open cluster candidates	Piatti et al. (2011)
VVV	Vista Variables in the Via Lactea	Minniti et al. (2010)
OGLE	Optical Gravitational Lensing Experiment	Udalski et al. (2008)
2MASS	Two Micron All Sky Survey	Skrutskie et al. (2006)
SAGE	Spitzer Survey of the Large Magellanic Cloud: Surveying the Agents of a Galaxy’s Evolution	Meixner et al. (2006)
CoRoT	Convection, Rotation, and planetary Transits	Baglin et al. (2006)
SDSS	Sloan Digital Sky Survey	York et al. (2000)
MACHO	Massive Compact Halo Objects survey	Alcock et al. (2000)
EROS	Expérience pour la Recherche d’Objets Sombres	Palanque-Delabrouille et al. (1998)
ASAS	All Sky Automated Survey	Pojmanski (1997)

Table 9
Synthetic Data Sources Used for ML Classification

Abbreviation	Long Name/Description	Reference
PLAsTiCC	Photometric LSST Astronomical Time-Series Classification Challenge	Kessler et al. (2019)
SNANA	SuperNova ANALysis software	Kessler et al. (2009)
SPCC	Supernova Photometric Classification Challenge	Kessler et al. (2010)
	Type II SN confined wind acceleration model	Moriya et al. (2019)
	Type Ia SN spectral templates	Hsiao et al. (2007)

Table 10
Pulsating Variable Star Classes (Excluding Red Giants and Supergiants) Found in the ML Literature (See Text for Further Details)

Type	Class Abbrev.	Brief Description
Lower MS	DSCT	δ Scutis. Low-order p -mode pulsators. Both radial and nonradial modes can be present. Periods typically shorter than 0.42 day. Pop. I.
	LBoo	λ Böotis. A-type MS dwarf with low metallicities. Part of the DSCT class.
	SXPh	SX Phoenixis. Pop. II counterparts of the DSCT. Typically found in globular clusters and dSph galaxies. Includes pulsating blue straggler stars.
	roAp	Rapidly oscillating Ap stars. High-order, nonradial p -mode pulsators. Amplitudes typically do not exceed 0.012 mag in V .
	GDor	γ Doradus. High-order, nonradial g -mode pulsators. Periods between 0.3 and 3 days, amplitudes less than 0.1 mag in V .
Upper MS	BCep	β Cepheids. Nonradial p -mode pulsators. Periods between 0.1 and 0.6 day, amplitudes in V between 0.01 and 0.32 mag.
	SPB	Slowly pulsating blue stars, aka 53 Per stars. Nonradial g -mode pulsators. Periods between 0.4–6 days, amplitudes in V less than 0.03 mag.
RR Lyrae	RRL(ab, c, d, Ad, e, GB)	RR Lyrae. Pulsating horizontal-branch stars with periods of order 0.5 day. Subtypes: ab (fundamental mode), c (first overtone), d (double mode), Ad (anomalous double mode), e (second overtone). Also classified by location (Galactic bulge, GB).
	Blazhko	RRL with long-period modulations (Blazhko effect).
Cepheids	Ceph(CL, F, O1, DM, MM, other)	δ Cepheids, aka classical (CL) or type I Cepheids. Pulsating G–K giant and supergiant stars. Often found pulsating in the fundamental (F), first (OI), or second overtone; double (DM) or multimode (MM) pulsation also common.
	ACEP	Anomalous Cepheids, aka BL Boo stars. Evolved counterparts of the SX Phe stars. Commonly found in dSph galaxies.
	CephII	

Table 10
(Continued)

Type	Class Abbrev.	Brief Description
	RVTau	Type II Cepheids. Low-mass Pop. II stars, often subdivided into BL Her, W Vir, and RV Tau subclasses with increasing periods. Type II Cepheids with periods in excess of 30 days. Light curves are well behaved and show double minima at the short-period end but become increasingly irregular with increasing period.
Subdwarf	sdBV	Pulsating subdwarf B stars, aka V361 Hya, EC 14026, sdBV _p , or sdBV _r stars; <i>p</i> -mode pulsators in which both radial and nonradial modes can be present. Periods between 60 and 570 s, amplitudes in <i>V</i> less than 65 mmag.
Compact	GW Vir	Pulsating pre-WD stars, aka pulsating PG 1159 stars. Includes both pulsating O-type WD stars and so-called planetary nebulae nucleus variables.
	DAV	Pulsating A-type WD stars, aka ZZ Ceti variables. Nonradial <i>g</i> -mode pulsators with H-dominated atmospheres.
	DBV	Pulsating B-type WD stars, aka V777 Her stars. Nonradial <i>g</i> -mode pulsators with He-dominated atmospheres.

Table 11
Same as Table 10 but for Pulsating Red Giants and Supergiants

Type	Class Abbrev.	Brief Description
Red giants	LPV	Long-period variable. Pulsating cool giant or supergiant stars. Often subdivided into Miras, SRs, Irregulars, and OSARGs.
	Mira	Mira variables. LPV red giants with very red colors and large amplitudes (by definition, exceeding 2.5 mag in <i>V</i>). Can be C- or O-rich, depending on evolutionary history.
	SR	Semiregular variables. Similar to the Miras but with smaller amplitudes (by definition, not exceeding 2.5 mag in <i>V</i>). Often subdivided into SRA (persistent periodicity), SRb (poorly defined periodicity), SRc (red supergiant SRs), and SRd (orange/yellow supergiant SRs).
	OSARG	OGLE small-amplitude red giant. Less evolved/luminous counterpart of the Miras and SRs, with smaller amplitudes and frequently multiple pulsation modes present.
	LPVW(A, B, C, D)	LPVs classified according to the sequence that they follow in a so-called Wood diagram (Wood et al. 1999).
	LPV(MAGB[C, O])	C- or O-rich Mira-type LPVs on the asymptotic giant branch (AGB)
	LPV(OSARGAGB)	OSARG-type LPVs on the AGB
	LPV(OSARGRGB[O])	Normal or O-rich OSARG-type LPVs on the red giant branch
	LPV(SRAGB[C, O])	C- or O-rich SR-type LPVs on the AGB
Supergiants	RSG	Red supergiant stars with irregular or semiregular light curves (Lc and SRc, respectively, as per the GCVS). According to Chatys et al. (2019), periodicities may include two groups related to pulsations ($P \sim 300\text{--}1000$ days) and LSPs ($P \sim 1000\text{--}8000$ days).
	LSP	LPV red giants with long secondary periods.
	PVSG	Periodic variable supergiant star.

Table 12
Stellar Variability Classes, Other than the Pulsating Ones, in the ML Literature (See Text for Further Details)

Var. Type	Class	Brief Description
Nonvariable	NV	Nonvariable star
Eclipsing	E(C, SD, D)	Eclipsing binary, classified according to its physical status as contact (C), semidetached (SD), or detached (D)
	BPer, BLyr, WUMa	Eclipsing binary, phenomenologically classified according to its light-curve shape into β Per (Algol, EA), β Lyr (EB), and W UMa (EW).
Rotational	ROT	Rotational variable. Rotating stars with nonuniform surface (starspots).
	ChemPec	Chemically peculiar rotational variable star.
	ELL	Close binary systems with ellipsoidal components (not eclipsing).
	RSCVn	RS Canum Venaticorum variable. Binary systems in which the primary star is typically a giant, characterized by semiperiodic light curves due to active chromospheres and the presence of starspots.
Chromosph.	ACT	Stars presenting surface activity due to active coronae and chromospheres.
	M dwarf	M dwarf flaring star; flares are caused by magnetic field reconnection events.
	[C, WL]TTau	Classic (C) or weak-lined (WL) T Tauri stars. Low-mass YSOs undergoing accretion from their surrounding disks. Depending on the H α emission strength, they are subdivided into C (strong emission) and WL (weak emission). Possible evolutionary link with EX Lupi (EXor) and FU Ori (FUor) stars, according to the mass accretion rate.
YSO	HAeBe	Herbig Ae/Be star. Higher-mass counterparts of the T Tauri stars. When large, irregular dust obscuration events are present, they may also be classified as UX Ori (UXor) stars.
	FUOri	FU Orionis stars. Pre-MS stars undergoing abrupt mass accretion episodes.

Table 12
(Continued)

Var. Type	Class	Brief Description
Outburst	LBV	Luminous blue variable (aka S Doradus) star. Hot, luminous stars near or above the Eddington limit undergoing vigorous mass loss and outbursts, followed by quiescent states.
	CV/nova	Cataclysmic variable star (including classical novae). Mass-transferring binary system in which an MS star transfers mass onto a WD via Roche lobe overflow. In the case of classical novae, thermonuclear explosions take place at the surface of the mass-accreting WD, followed by a quiescent state.
Lensing	ML	Microlensing event. Star whose brightness is magnified due to a gravitational lensing event.
Other	RCB	R Coronae Borealis stars. F- or G-type self-eclipsing supergiant stars that undergo dramatic dimming events brought about by mass-loss episodes followed by dust condensation.
	DPV	Double periodic variable. Binary system with variability due to eclipses or ellipsoidal modulations on timescales of order a few days, accompanied by a long cycle lasting about 33 times the orbital period.
	BeS	Be stars. Nonsupergiant B star rotating close to breakup speed and presenting decretion disks, accompanied by variable Balmer emission.
	LAPV	Low-amplitude periodic variable. Defined in Debosscher et al. (2009), including low-amplitude Cepheids and also rotational variable stars with regular light curves.
	WR	Wolf-Rayet star. Evolved, massive stars that have lost their H envelopes and show signatures of strong stellar winds.
	XB	X-ray binary. CV-like systems in which the accreting star is typically not a WD but rather a neutron star or BH and which thus emit their energy mostly in the form of X-rays.

Table 13
Extragalactic BH-related Variability Classes as Found in the ML Literature (See Text for Further Details)

Abbreviation	Description
AGN	Active galactic nuclei. Central accreting SMBH ($>10^5 M_\odot$) where the host galaxy dominates the total light. Variability likely due to accretion-disk instabilities.
QSO	Quasi-stellar object. Central accreting SMBH that dominates over the host galaxy in the total light. Variability likely due to accretion disk instabilities.
Blazar	Central accreting SMBH with a relativistic jet directed toward the observer. Variability due to synchrotron and inverse Compton relativistic beaming. This category does not distinguish between blazars, BL Lacs, and optical violent variables, which peak in different wave bands.

Table 14
Transient Classes as Found in the ML Literature (See Text for Further Details)

Abbreviation	Description
SN Ia	Type Ia SNe. Thermonuclear explosion of a CO white dwarf.
SN Ia-91bg	Underluminous SNe Ia. SN 1991bg-like.
SN Iax	Type Iax SNe. Deflagration-dominated SNe Ia.
.Ia	“Ia” SNe. He shell detonation explosion.
SN Ibc	Type Ib or Ic SNe. Core collapse (CC) of envelope-stripped massive star.
SN II	Type II SNe. CC of red supergiant star.
SN IIn	Type IIn SNe. SN explosion in dense circumstellar medium.
TDE	Tidal disruption event. Stellar disruption due to BH proximity.
CART	Calcium-rich transient.
ILOT	Intermediate-luminosity optical transient.
PISN	Pair instability SNe. CC and thermonuclear explosion due to e^-/e^+ pair production.
SLSN	Superluminous SNe. Class of explosions about 10 times brighter than standard SNe.
kN	Kilonova. Neutron star merger optical counterpart.

In the case of Tables 10 and 11, the pulsating variable star classes are shown. Table 10 includes pulsating stars in the upper and lower main sequence, Cepheids, RR Lyrae, blue subdwarfs, and compact (WD) pulsators. Table 11, in turn, includes red giant and supergiant pulsators.



Table 12 presents a number of additional stellar variability classes, including eclipsing, eruptive, cataclysmic, and rotational variables. Additional classes that are shown in this table

include microlensing events, R CrB stars, Be stars, and X-ray binaries, among others.

Primarily extragalactic variable sources are shown in Tables 13 and 14. In the case of Table 13, the variability is typically related to the presence of SMBHs, as in the case of AGNs and QSOs. Table 14, in turn, primarily includes a variety of SN classes, although a few transient events of non-SN origin, such as TDEs and kilonovae, are also included.

We emphasize that the classes and associated taxonomies that are implied by Tables 5–14 do not reflect our own choices but rather are simply a summary of what has been used in the ML literature to date. In particular, the reader should be aware that the list of classes as given suffers from several shortcomings, such as being incomplete, containing redundant entries, and including classes that may not be sufficiently well defined. Still, our best effort to interpret what the different authors have intended to express in each case is reflected in these tables, with definitions given following, among others, the General Catalog of Variable Stars (GCVS; Kholopov et al. 1998), the Variable Star Index (Watson et al. 2006), and the broad overview of stellar variability classes presented in Catelan & Smith (2015). In the future, as the ALerCE project matures, we will work toward producing and refining our own taxonomy, which we will perfect along the way as we enter the LSST era.

ORCID iDs

F. Förster  <https://orcid.org/0000-0003-3459-2270>
 P. A. Estévez  <https://orcid.org/0000-0001-9164-4722>
 P. Sánchez-Sáez  <https://orcid.org/0000-0003-0820-4692>
 F. E. Bauer  <https://orcid.org/0000-0002-8686-8737>
 M. Catelan  <https://orcid.org/0000-0001-6003-8877>
 F. Elorrieta  <https://orcid.org/0000-0002-1835-7433>
 P. Huijse  <https://orcid.org/0000-0003-3541-1697>
 G. Pignata  <https://orcid.org/0000-0003-0006-0188>
 E. Reyes  <https://orcid.org/0000-0003-3627-0216>
 C. Valenzuela  <https://orcid.org/0000-0001-5306-1390>
 J. Borissova  <https://orcid.org/0000-0002-5936-7718>
 L. Hernández-García  <https://orcid.org/0000-0002-8606-6961>
 M. J. Graham  <https://orcid.org/0000-0002-3168-0139>
 A. Jordán  <https://orcid.org/0000-0002-5389-3944>
 R. Kurtev  <https://orcid.org/0000-0002-9740-9974>
 A. Mahabal  <https://orcid.org/0000-0003-2242-0244>
 J. C. Maureira  <https://orcid.org/0000-0002-7458-6142>
 A. Muñoz-Arancibia  <https://orcid.org/0000-0002-8722-516X>
 M. Pérez-Carrasco  <https://orcid.org/0000-0003-4644-8698>
 A. Sánchez  <https://orcid.org/0000-0003-0820-4692>
 J. R. Vergara  <https://orcid.org/0000-0001-6699-4181>

References

- Aartsen, M. G., Ackermann, M., Adams, J., et al. 2017, *JInst*, **12**, P03012
 Abadi, M., Barham, P., Chen, J., et al. 2016, in Proc. 12th USENIX Symp. on Operating Systems Design and Implementation (OSDI '16). (Berkeley, CA: USENIX), 265, <https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf>
 Abbasi, R., Ackermann, M., Adams, J., et al. 2009, *NIMPA*, **601**, 294
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *ApJL*, **848**, L12
 Abramovici, A., Althouse, W. E., Drever, R. W. P., et al. 1992, *Sci*, **256**, 325
 Abril, J., Schmidtobreick, L., Ederoclite, A. r., & López-Sanjuan, C. 2020, *MNRAS*, **492**, L40
 Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, *CQGra*, **32**, 024001
 Actis, M., Agnetta, G., Aharonian, F., et al. 2011, *ExA*, **32**, 193
 Aguirre, C., Pichara, K., & Becker, I. 2019, *MNRAS*, **482**, 5078
 Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, *PASJ*, **70**, S4
 Alcock, C., Allsman, R. A., Alves, D. R., et al. 2000, *ApJ*, **542**, 281
 Alexander, D. M., & Hickox, R. C. 2012, *NewAR*, **56**, 93
 Andreoni, I., Cooke, J., Webb, S., et al. 2020, *MNRAS*, **491**, 5852
 Arcavi, I., Gal-Yam, A., Sullivan, M., et al. 2014, *ApJ*, **793**, 38
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
 Atwood, W. B., Abdo, A. A., Ackermann, M., et al. 2009, *ApJ*, **697**, 1071
 Baglin, A., Auvergne, M., Boissard, L., et al. 2006, *COSPAR*, **36**, 3749
 Bailey, S., Aragon, C., Romano, R., et al. 2007, *ApJ*, **665**, 1246
 Baron, D., & Poznanski, D. 2017, *MNRAS*, **465**, 4530
 Bass, G., & Borne, K. 2016, *MNRAS*, **459**, 3721
 Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, *PASP*, **131**, 018002
 Benavente, P., Protopapas, P., & Pichara, K. 2017, *ApJ*, **845**, 147
 Bennett, D. P., & Rhie, S. H. 1996, *ApJ*, **472**, 660
 Bensby, T., Feltzing, S., Yee, J. C., et al. 2020, *A&A*, **634**, A130
 Bloom, J. S., Richards, J. W., Nugent, P. E., et al. 2012, *PASP*, **124**, 1175
 Boch, T., & Desrozières, J. 2020, in ASP Conf. Ser., 522, *Astronomical Data Analysis Software and Systems XXVII*, ed. P. Ballester et al. (San Francisco, CA: ASP), 117
 Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *A&AS*, **143**, 33
 Boone, K. 2019, *AJ*, **158**, 257
 Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Sci*, **327**, 977
 Boulade, O., Charlot, X., Abbon, P., et al. 2003, *Proc. SPIE*, **4841**, 72
 Brahm, M., Yeh, L., & Zeitouni, K. 2018, in SIGSPATIAL '18: Proc. 26th ACM SIGSPATIAL Int. Conf. Advances in Geographic Information Systems, ed. F. Banaei-Kashani & E. Hoel (New York: ACM), 229
 Brunel, A., Pasquet, J., Pasquet, J., et al. 2019, arXiv:1901.00461
 Buchler, J. R., & Kolláth, Z. 2002, *ApJ*, **573**, 324
 Cabrera-Vives, G., Reyes, I., Förster, F., Estévez, P. A., & Maureira, J. C. 2017, *ApJ*, **836**, 97
 Cappellaro, E. 2005, *Msngr*, **120**, 13
 Carrasco, D., Barrientos, L. F., Pichara, K., et al. 2015, *A&A*, **584**, A44
 Carrasco-Davis, R., Cabrera-Vives, G., Förster, F., et al. 2019, *PASP*, **131**, 108006
 Carrasco-Davis, R., Reyes, E., Valenzuela, C., et al. 2020, arXiv:2008.03309
 Castro, N., Protopapas, P., & Pichara, K. 2018, *AJ*, **155**, 16
 Catelan, M., & Smith, H. A. 2015, *Pulsating Stars* (Weinheim: Wiley & Sons)
 Chapelle, O., Scholkopf, B., & Zien, A. 2009, *ITNN*, **20**, 542
 Charnock, T., & Moss, A. 2017, *ApJL*, **837**, L28
 Chatys, F. W., Bedding, T. R., Murphy, S. J., et al. 2019, *MNRAS*, **487**, 4832
 Clement, C. M., & Goranskij, V. P. 1999, *ApJ*, **513**, 767
 Connelley, M. S., & Reipurth, B. 2018, *ApJ*, **861**, 145
 Contreras Peña, C., Lucas, P. W., Minniti, D., et al. 2017, *MNRAS*, **465**, 3011
 Dalton, G. B., Caldwell, M., Ward, A. K., et al. 2006, *Proc. SPIE*, **6269**, 62690X
 Dau, H. A., Keogh, E., Kamgar, K., et al. 2018, arXiv:1810.07758
 Debosscher, J., Sarro, L. M., Aerts, C., et al. 2007, *A&A*, **475**, 1159
 Debosscher, J., Sarro, L. M., López, M., et al. 2009, *A&A*, **506**, 519
 Dékány, I., Hajdu, G., Grebel, E. K., & Catelan, M. 2019, *ApJ*, **883**, 58
 Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, *IEEEP*, **97**, 1482
 D'Isanto, A., Cavuoti, S., Brescia, M., et al. 2016, *MNRAS*, **457**, 3119
 Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, *ApJ*, **696**, 870
 Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, *ApJS*, **213**, 9
 Drake, A. J., Djorgovski, S. G., Catelan, M., et al. 2017, *MNRAS*, **469**, 3688
 Drout, M. R., Chornock, R., Soderberg, A. M., et al. 2014, *ApJ*, **794**, 23
 du Buisson, L., Sivanandam, N., Bassett, B. A., & Smith, M. 2015, *MNRAS*, **454**, 2026
 Duev, D. A., Mahabal, A., Masci, F. J., et al. 2019, *MNRAS*, **489**, 3582
 Elorrieta, F., Eyheramendy, S., & Palma, W. 2019, *A&A*, **627**, A120
 Estévez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M. 2009, *ITNN*, **20**, 189
 Eyheramendy, S., Elorrieta, F., & Palma, W. 2018, *MNRAS*, **481**, 4311
 Fang, Q., Maeda, K., Kuncarayakti, H., Sun, F., & Gal-Yam, A. 2019, *NatAs*, **3**, 434
 Faraway, J., Mahabal, A., Sun, J., et al. 2016, *SADM*, **9**, 1
 Faria, E. R., Gonçalves, I. J., de Carvalho, A. C., & Gama, J. 2016, *Artif Intell Rev*, **45**, 235
 Fernique, P., Allen, M. G., Boch, T., et al. 2015, *A&A*, **578**, A114
 Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, **150**, 150
 Flesch, E. W. 2019, arXiv:1912.05614
 Förster, F., Maureira, J. C., San Martín, J., et al. 2016, *ApJ*, **832**, 155
 Förster, F., Moriya, T. J., Maureira, J. C., et al. 2018, *NatAs*, **2**, 808
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, **616**, A1
 Gaia Collaboration, Eyer, L., Rimoldini, L., et al. 2019a, *A&A*, **623**, A110
 Gaia Collaboration, Eyer, L., Rimoldini, L., et al. 2019b, *A&A*, **623**, A110
 Gal-Yam, A., Arcavi, I., Ofek, E. O., et al. 2014, *Natur*, **509**, 471
 Gould, A., Dong, S., Gaudi, B. S., et al. 2010, *ApJ*, **720**, 1073
 Graham, M. J., Ross, N. P., Stern, D., et al. 2020, *MNRAS*, **491**, 4925
 Green, A. M. 2016, *PhRvD*, **94**, 063530
 Greene, J. E., Strader, J., & Ho, L. C. 2020, *ARA&A*, **58**, 257
 Guillochon, J., Parent, J., Kelley, L. Z., & Margutti, R. 2017, *ApJ*, **835**, 64

- Gulli, A., & Pal, A. 2017, Deep Learning with Keras (Birmingham: Packt)
- Hosenie, Z., Lyon, R. J., Stappers, B. W., & Mootoovallu, A. 2019, *MNRAS*, **488**, 4858
- Hsiao, E. Y., Conley, A., Howell, D. A., et al. 2007, *ApJ*, **663**, 1187
- Huber, M., Rest, A., Narayan, G., et al. 2011, AAS Meeting, **218**, 328.12
- Huenteleu, P., Estévez, P. A., & Förster, F. 2016, in 2016 Int. Joint Conf. on Neural Networks (IJCNN) (Piscataway, NJ: IEEE), 3322
- Huijse, P., Estevez, P. A., Förster, F., et al. 2018, *ApJS*, **236**, 12
- Hunter, J. D. 2007, *CSE*, **9**, 90
- IceCube Collaboration, Aartsen, M. G., Ackermann, M., et al. 2018, *Sci*, **361**, 147
- Jacyszyn-Dobrzaniecka, A. M., Soszyński, I., Udalski, A., et al. 2020a, *ApJ*, **889**, 25
- Jacyszyn-Dobrzaniecka, A. M., Mróz, P., Kruszyńska, K., et al. 2020b, *ApJ*, **889**, 26
- Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, *MNRAS*, **486**, 1907
- Johnston, K. B., Caballero-Nieves, S. M., Petit, V., Peter, A. M., & Haber, R. 2020, *MNRAS*, **491**, 3805
- Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, *Proc. SPIE*, **4836**, 154
- Karpenka, N. V., Feroz, F., & Hobson, M. P. 2013, *MNRAS*, **429**, 1278
- Kasliwal, M. M., Kulkarni, S. R., Gal-Yam, A., et al. 2012, *ApJ*, **755**, 161
- Keller, S. C., Schmidt, B. P., Bessell, M. S., et al. 2007, *PASA*, **24**, 1
- Kessler, R., Conley, A., Jha, S., & Kuhlmann, S. 2010, arXiv:1001.5210
- Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, *PASP*, **121**, 1028
- Kessler, R., Narayan, G., Avelino, A., et al. 2019, *PASP*, **131**, 094501
- Kholopov, P. N., Samus, N. N., Frolov, M. S., et al. 1998, General Catalogue of Variable Stars, Vol. I-III (4th ed.; Moscow: Nauka Publishing House)
- Kim, D. W., & Bailer-Jones, C. A. L. 2016, *A&A*, **587**, A18
- Kim, D.-W., Protopapas, P., Bailer-Jones, C. A., et al. 2014, *A&A*, **566**, A43
- Kim, S.-L., Lee, C.-U., Park, B.-G., et al. 2016, *JKAS*, **49**, 37
- Kimura, A., Takahashi, I., Tanaka, M., et al. 2017, in IEEE 37th Int. Conf. on Distributed Computing Systems Workshops (ICDCSW), ed. A. Musaeu, J. E. Ferreira, & T. Higashino (Piscataway, NJ: IEEE), 354
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt (Amsterdam: IOS Press), 90
- Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *ApJL*, **713**, L79
- Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, *PASP*, **129**, 104502
- Komossa, S. 2015, *JHEAp*, **7**, 148
- Kügler, S. D., Gianniotis, N., & Polsterer, K. L. 2015, *MNRAS*, **451**, 3385
- LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, *ApJ*, **800**, 144
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, *PASP*, **121**, 1395
- Law, N. M., Fors, O., Ratzloff, J., et al. 2015, *PASP*, **127**, 234
- Lawrence, A. 2016, in ASP Conf. Ser. Vol. 505, Astronomical Surveys and Big Data: Proceedings of an International Symposium dedicated to the 50th anniversary of the Markarian Survey and the 10th anniversary of the Armenian Virtual Observatory, ed. A. M. Mickaelian, A. Lawrence, & T. Yu. Magakian (San Francisco, CA: ASP), 107
- Lemaître, G., Nogueira, F., & Aridas, C. K. 2017, JMLR, **18**, 1, <https://jmlr.org/papers/volume18/16-365/16-365.pdf>
- Liu, W., Principe, J. C., & Haykin, S. 2011, Kernel Adaptive Filtering: A Comprehensive Introduction (Hoboken, NJ: John Wiley & Sons)
- Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, *ApJS*, **225**, 31
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
- Mackenzie, C., Pichara, K., & Protopapas, P. 2016, *ApJ*, **820**, 138
- Mahabal, A., Sheth, K., Gieseke, F., et al. 2017, in IEEE Symp. Ser. on Computational Intelligence (SSCI) (Piscataway, NJ: IEEE), 1
- Martínez-Palomera, J., Lira, P., Bhalla-Ladd, I., Förster, F., & Plotkin, R. M. 2020, *ApJ*, **889**, 113
- Martínez-Palomera, J., Förster, F., Protopapas, P., et al. 2018, *AJ*, **156**, 186
- Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, *PASP*, **131**, 018003
- Massaro, E., Maselli, A., Leto, C., et al. 2015, *Ap&SS*, **357**, 75
- McKinney, W., et al. 2010, in Proc. 9th Python in Science Conference, ed. S. van der Walt & J. Millman (Austin, TX: SciPy), 56
- McLeod, B., Geary, J., Conroy, M., et al. 2015, *PASP*, **127**, 366
- Meixner, M., Gordon, K. D., Indebetouw, R., et al. 2006, *AJ*, **132**, 2268
- Merloni, A., Predehl, P., Becker, W., et al. 2012, arXiv:1209.3114
- Mezcua, M. 2017, *IJMPD*, **26**, 1730021
- Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, *NewA*, **15**, 433
- Möller, A., & de Boissière, T. 2020, *MNRAS*, **491**, 4277
- Möller, A., Peloton, J., Ishida, E. E. O., et al. 2021, *MNRAS*, **501**, 3272
- Moriya, T. J., Liu, D., Wang, B., & Liu, Z.-W. 2019, *MNRAS*, **488**, 3949
- Morokuma, T., Tominaga, N., Tanaka, M., et al. 2014, *PASJ*, **66**, 114
- Mowlavi, N., Lecoœur-Taïbi, I., Lebzelter, T., et al. 2018, *A&A*, **618**, A58
- Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, *PASP*, **131**, 118002
- Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, *ApJS*, **236**, 9
- Naul, B., Bloom, J. S., Pérez, F., & van der Walt, S. 2018, *NatAs*, **2**, 151
- Nordin, J., Brinnel, V., van Santen, J., et al. 2019, *A&A*, **631**, A147
- Nun, I., Protopapas, P., Sim, B., & Chen, W. 2016, *AJ*, **152**, 71
- Nun, I., Protopapas, P., Sim, B., et al. 2017, arXiv:1506.00010
- Oh, K., Yi, S. K., Schawinski, K., et al. 2015, *ApJS*, **219**, 1
- Palanque-Delabrouille, N., Afonso, C., Albert, J. N., et al. 1998, *A&A*, **332**, 1
- Palaversa, L., Ivezić, Ž., Eyer, L., et al. 2013, *AJ*, **146**, 101
- Patterson, M. T., Bellm, E. C., Rusholme, B., et al. 2019, *PASP*, **131**, 018001
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, **12**, 2825, <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, *ApJ*, **613**, 682
- Piatti, A. E., Clariá, J. J., & Ahumada, A. V. 2011, *NewA*, **16**, 161
- Pichara, K., & Protopapas, P. 2013, *ApJ*, **777**, 83
- Pichara, K., Protopapas, P., & Leo'n, D. 2016, *ApJ*, **819**, 18
- Pietrzyński, G., Thompson, I. B., Gieren, W., et al. 2010, *Natur*, **468**, 542
- Pietrzyński, G., Thompson, I. B., Gieren, W., et al. 2012, *Natur*, **484**, 75
- Pojmanski, G. 1997, *AcA*, **47**, 467
- Revsbech, E. A., Trotta, R., & van Dyk, D. A. 2018, *MNRAS*, **473**, 3969
- Reyes, E., & Estévez, P. A. 2020, arXiv:2005.07779
- Richards, J. W., Starr, D. L., Miller, A. A., et al. 2012, *ApJS*, **203**, 32
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, *JATIS*, **1**, 014003
- Rimoldini, L., Holl, B., Audard, M., et al. 2019, *A&A*, **625**, A97
- Ritter, H., & Kolb, U. 2003, *A&A*, **404**, 301
- Rocklin, M. 2015, in Proc. 14th Python in Science Conference, ed. K. Huff & J. Bergstra (Austin, TX: SciPy), 130
- Ross, N. P., Ford, K. E. S., Graham, M., et al. 2018, *MNRAS*, **480**, 4468
- Sánchez-Sáez, P., Lira, P., Mejía-Restrepo, J., et al. 2018, *ApJ*, **864**, 87
- Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021, *AJ*, **161**, 141
- Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., & Williamson, R. 2001, *Neural Computation*, **13**, 1443
- Sesar, B., Stuart, J. S., Ivezić, Ž., et al. 2011, *AJ*, **142**, 190
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Smith, K. W., Williams, R. D., Young, D. R., et al. 2019, *RNAAS*, **3**, 26
- Soszyński, I., Dziembowski, W. A., Udalski, A., et al. 2014, *AcA*, **64**, 1
- Soumagnac, M. T., & Ofek, E. O. 2018, *PASP*, **130**, 075002
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
- Steer, I., Madore, B. F., Mazzarella, J. M., et al. 2017, *AJ*, **153**, 37
- Street, R. A., Bowman, M., Saunders, E. S., & Boroson, T. 2018, *Proc. SPIE*, **10707**, 1070711
- The Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0510346
- Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, *PASP*, **130**, 064505
- Tsang, B. T. H., & Schultz, W. C. 2019, *ApJL*, **877**, L14
- Tyson, J. A. 2019, *JPhCS*, **1290**, 012001
- Udalski, A., Szymanski, M., Soszynski, I., & Poleski, R. 2008, *AcA*, **58**, 69
- Udalski, A., Szymański, M. K., & Szymański, G. 2015, *AcA*, **65**, 1
- Valenzuela, L., & Pichara, K. 2018, *MNRAS*, **474**, 3259
- Van Rossum, G., & Drake, F. L., Jr. 1995, Python Reference Manual (Amsterdam: Centrum Wiskunde & Informatica)
- Villar, V. A., Berger, E., Miller, G., et al. 2019, *ApJ*, **884**, 83
- Vivas, A. K., Martínez-Vázquez, C., & Walker, A. R. 2020, *ApJS*, **247**, 35
- Vivas, A. K., Zinn, R., Abad, C., et al. 2004, *AJ*, **127**, 1158
- Watson, C. L., Henden, A. A., & Price, A. 2006, *SASS*, **25**, 47
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, **143**, 9
- Wood, P. R., Alcock, C., Allsman, R. A., et al. 1999, in IAU Symposium, Vol. 191, Asymptotic Giant Branch Stars, ed. T. Le Bertre, A. Lebre, & C. Waelkens (San Francisco, CA: ASP), 151
- Wright, D. E., Lintott, C. J., Smartt, S. J., et al. 2017, *MNRAS*, **472**, 1315
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868
- Yaron, O., Perley, D. A., Gal-Yam, A., et al. 2017, *NatPh*, **13**, 510
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Zaharia, M., Xin, R. S., Wendell, P., et al. 2016, *CACM*, **59**, 56
- Zhu, X., Ghahramani, Z., & Lafferty, J. 2003, in Proc. Twentieth Int. Conf. on Machine Learning (ICML 2003), ed. T. Fawcett & N. Mishra (Menlo Park, CA: AAAI Press), 912, <https://aaai.org/Library/ICML/2003/icml03-118.php>
- Zinn, J. C., Kochanek, C. S., Kozłowski, S., et al. 2017, *MNRAS*, **468**, 2189